B146

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 February 2001 (22.02.2001)

PCT

(10) International Publication Number
WO 01/12790 A2

(51) International Patent Classification[7]: C12N 9/00

(21) International Application Number: PCT/US00/22518

(22) International Filing Date: 16 August 2000 (16.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/149,388 17 August 1999 (17.08.1999) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/149,388 (CIP)
Filed on 17 August 1999 (17.08.1999)

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BANDMAN, Olga** [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). **LU, Dyung, Aina, M.** [US/US]; 55 Park Belmont Place, San Jose, CA 95136 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **TRAN, Bao** [US/US]; 750 Salberg Avenue, Santa Clara, CA 95051 (US). **HILLMAN, Jennifer, L.** [US/US]; 230 Monroe Drive, #17, Mountain View, CA 94040 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **LAL, Preeti** [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). **TANG, Y., Tom** [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/12790 A2

(54) Title: ISOMERASE PROTEINS

(57) Abstract: The invention provides human isomerases (ISOM) and polynucleotides which identify and encode ISOM. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of ISOM.

ISOMERASE PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of isomerases and to the use of these sequences in the diagnosis, treatment, and prevention of immune and cell proliferation disorders including cancer.

BACKGROUND OF THE INVENTION

Proteins may be modified after translation through structural rearrangements and/or by the addition of phosphate, sugar, prenyl, fatty acid, and other chemical groups. Cells contain a number of specialized molecules that assist in the formation of protein structure. Enzymes involved in post-translational modification include kinases, phosphatases, glycosyltransferases, and prenyltransferases. These molecules facilitate the folding of newly synthesized proteins, prevent aggregation and improper glycosylation, and remove denatured protein. These modifications are often required for proper protein activity. The conformation of proteins may also be modified after translation for example, by the introduction and rearrangement of disulfide bonds (rearrangement catalyzed by protein disulfide isomerases), the isomerization of proline side chains by prolyl isomerase, and by interactions with molecular chaperone proteins.

Numerous essential biochemical reactions involve the isomerization of a substrate. Enzymes which catalyze such reactions are known as isomerases. Isomerases are a class of enzymes that catalyze geometric or structural changes within a molecule to form a single product. A number of isomerases have been described catalyzing steps in a wide variety of biochemical pathways including protein folding, phototransduction, and various anabolic and catabolic pathways (e.g., glycolysis), in organisms ranging from bacteria to humans. Within the class of isomerases are racemases and epimerases, *cis-trans*-isomerases, intramolecular oxidoreductases and intramolecular transferases (mutases).

Racemases and Epimerases

Racemases are a subset of isomerases that catalyze inversion of a molecules configuration around the asymmetric carbon atom in a substrate having a single center of asymmetry, thereby interconverting two racemers. Epimerases are another subset of isomerases that catalyze inversion of configuration around an asymmetric carbon atom in a substrate with more than one center of symmetry, thereby interconverting two epimers. Racemases and epimerases can act on amino acids and derivatives, hydroxy acids and derivatives, as well as carbohydrates and derivatives. The interconversion of UDP-galactose and UDP-glucose is catalyzed by UDP-galactose-4'-epimerase. Proper regulation and function of this epimerase is essential to the synthesis of glycoproteins and

glycolipids. Elevated blood galactose levels have been correlated with UDP-galactose-4'-epimerase deficiency in screening programs of infants (Gitzelmann, R. (1972) Helv. Paediat. Acta 27:125-130).

Peptidyl prolyl *cis-trans* isomerases

Correct folding of newly synthesized proteins is assisted by molecular chaperones and folding catalysts, two unrelated groups of helper molecules. Chaperones suppress non-productive side reactions by stoichiometric binding to folding intermediates, whereas folding enzymes catalyze some of the multiple folding steps that enable proteins to attain their final functional configurations (Kern, G. et al. (1994) FEBS Lett. 348: 145-148). One class of folding enzymes, the peptidyl prolyl *cis-trans* isomerases (PPIases), isomerize certain proline imidic bonds in what is considered to be a rate limiting step in protein maturation and export. PPIases catalyze the *cis* to *trans* isomerization of certain proline imidic bonds in proteins. There are three sequence-unrelated families of PPIases, the cyclophilins, the FK506 binding proteins, and the newly characterized parvulin family (Rahfeld, J.U. et al. (1994) FEBS Lett. 352: 180-184).

The cyclophilins (CyP), were originally identified as major receptors for the immunosuppressive drug cyclosporin A (CsA) an inhibitor of T-cell activation (Handschumacher, R.E. et al. (1984) Science 226: 544-547; Harding, M.W. et al. (1986) J. Biol. Chem. 261: 8547-8555). Thus, the peptidyl-prolyl isomerase activity of CyP may be part of the signaling pathway that leads to T-cell activation. Subsequent work demonstrated that CyP's isomerase activity is essential for correct protein folding and/or protein trafficking, and may also be involved in assembly/disassembly of protein complexes and regulation of protein activity. For example, in *Drosophila*, the CyP NinaA is required for correct localization of rhodopsins, while a mammalian CyP (Cyp40) is part of the Hsp90/Hsp70 complex that binds steroid receptors. The mammalian CyP (CypA) has been shown to bind the *gag* protein from human immunodeficiency virus 1 (HIV-1), an interaction that can be inhibited by cyclosporin. Since cyclosporin has potent anti-HIV-1 activity, CypA may play an essential function in HIV-1 replication. Finally, Cyp40 has been shown to bind and inactivate the transcription factor c-Myb, an effect that is reversed by cyclosporin. This effect implicates CyPs in the regulation of transcription, transformation, and differentiation (Bergsma, D.J. et al (1991) J. Biol. Chem. 266:23204 - 23214; Hunter, T. (1998) Cell 92: 141-143; and Leverson, J.D. and Ness, S.A. (1998) Mol. Cell. 1:203-211).

Protein Disulfide Isomerases

One of the major rate limiting steps in protein folding is the thiol:disulfide exchange that is necessary for correct protein assembly. Although incubation of reduced, unfolded proteins in buffers with defined ratios of oxidized and reduced thiols can lead to native conformation, the rate of folding is slow and the attainment of native conformation decreases proportionately to the size and number of cysteines in the protein. Certain cellular compartments such as the endoplasmic reticulum of

eukaryotes and the periplasmic space of prokaryotes are maintained in a more oxidized state than the surrounding cytosol. Correct disulfide formation can occur in these compartments but at a rate that is insufficient for normal cell processes and not adequate for synthesizing secreted proteins. The protein disulfide isomerases, thioredoxins and glutaredoxins are able to catalyze the formation of disulfide bonds and regulate the redox environment in cells to enable the necessary thiol:disulfide exchanges (Loferer, H. (1995) J. Biol Chem. 270:26178-26183).

Each of these proteins have somewhat different functions but all belong to a group of disulfide-containing redox proteins that contain a conserved active-site sequence and are ubiquitously distributed in eukaryotes and prokaryotes. Protein disulfide isomerases are found in the endoplasmic reticulum of eukaryotes and in the periplasmic space of prokaryotes. They function by exchanging their own disulfide for a thiol in a folding peptide chain. In contrast, the reduced thioredoxins and glutaredoxins are generally found in the cytoplasm and function by directly reducing disulfides in the substrate proteins.

These catalytic molecules not only facilitate disulfide formation but also regulate and participate in a wide variety of physiological processes. The thioredoxin system serves, for example, as a hydrogen donor for ribonucleotide reductase and as a regulator of enzymes by redox control. It also modulates the activity of transcription factors such as NF-κB, AP-1, and steroid receptors. More recently, several cytokines or secreted cytokine-like factors such as adult T-cell leukemia-derived factor, 3B6-interleukin-1, T-hybridoma-derived (MP-6) B cell stimulatory factor, and early pregnancy factor have been reported to be identical to thioredoxin (Holmgren, A. (1985) Annu. Rev. Biochem. 54:237-271, Abate, C. et al., (1990) Science 249:1157-1161, Tagaya, Y. et al. (1989) EMBO J. 8:757-764, Wakasugi, H. (1987) Proc. Natl. Acad. Sci. 84:804-808, Rosen, A. et al. (1995) Int. Immunol. 7:625-633). Thioredoxin has also been shown to have many extracellular activities including a role as a regulator of cell growth and a mediator in the immune system (Miranda-Vizuete, A. et al. (1996) J. Biol. Chem. 271:19099-19103, Yamauchi, A. et al (1992) Mol. Immunol. 29:263-270).

Intramolecular oxidoreductases

Oxidoreductases can be isomerases as well. Oxidoreductases catalyze the reversible transfer of electrons from a substrate that becomes oxidized to a substrate that becomes reduced. This class of enzymes includes dehydrogenases, hydroxylases, oxidases, oxygenases, peroxidases, and reductases. Proper maintenance of oxidoreductase levels is physiologically important. The pentose phosphate pathway for example, utilizes enzymes which are responsible for generating the reducing agent NADPH, while at the same time oxidizing glucose-6-phosphate to ribose-5-phosphate. NADPH serves as the fuel for reactions undergoing reductive biosynthesis. Ribose-5-phosphate and its derivatives become part of critical biological molecules such as ATP, Coenzyme A, NAD^+, FAD,

RNA, and DNA. The pentose phosphate pathway has both oxidative and non-oxidative branches. The oxidative branch steps, which are catalyzed by the enzymes glucose-6-phosphate dehydrogenase, lactonase, and 6-phosphogluconate dehydrogenase, convert glucose-6-phosphate and $NADP^+$ to ribulose-6-phosphate and NADPH. The non-oxidative branch steps, which are catalyzed by the enzymes phosphopentose isomerase, phosphopentose epimerase, transketolase, and transaldolase, allow the interconversion of three-, four-, five-, six-, and seven-carbon sugars.

Transferases

Another subgroup of isomerases are the transferases (or mutases). Transferases transfer a chemical group from one compound (the donor) to another compound (the acceptor). The types of groups transferred by these enzymes include acyl groups, amino groups, phosphate groups (phosphotransferases or phosphomutases), and others. The transferase carnitine palmitoyltransferase is an important component of fatty acid metabolism. Genetically-linked deficiencies in this transferase can lead to myopathy (Scriver C.R. et. al. (1995) The Metabolic and Molecular Basis of Inherited Disease, McGraw-Hill New York NY pp.1501-1533).

Isomerases are critical components of cellular biochemistry with roles in metabolic energy production including glycolysis, as well as other diverse enzymatic processes (Stryer, L. (1995) Biochemistry W.H. Freeman and Co. New York, NY pp.483-507).

The discovery of new isomerases and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of immune and cell proliferation disorders including cancer.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, isomerases, referred to collectively as "ISOM" and individually as "ISOM-1," "ISOM-2," "ISOM-3," "ISOM-4," "ISOM-5," "ISOM-6," "ISOM-7," and "ISOM-8." In one aspect, the invention provides an isolated polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-8.

The invention further provides an isolated polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a naturally occurring amino acid sequence having at

least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-8. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:9-16.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

The invention further provides an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group

consisting of SEQ ID NO:9-16, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and a pharmaceutically

acceptable excipient. In one embodiment, the composition comprises an amino acid sequence sel cted from the group consisting of SEQ ID NO:1-8. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional ISOM, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional ISOM, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional ISOM, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of an amino acid sequenc

selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:9-16, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, ii) a naturally

occurring polynucleotide sequenc having at least 90% sequenc identity to a polynucleotid sequence selected from the group consisting of SEQ ID NO:9-16, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 shows polypeptide and nucleotide sequence identification numbers (SEQ ID NOs), clone identification numbers (clone IDs), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding ISOM.

Table 2 shows features of each polypeptide sequence, including potential motifs, homologous sequences, and methods, algorithms, and searchable databases used for analysis of ISOM.

Table 3 shows selected fragments of each nucleic acid sequence; the tissue-specific expression patterns of each nucleic acid sequence as determined by northern analysis; diseases, disorders, or conditions associated with these tissues; and the vector into which each cDNA was cloned.

Table 4 describes the tissues used to construct the cDNA libraries from which cDNA clones encoding ISOM were isolated.

Table 5 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a

reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"ISOM" refers to the amino acid sequences of substantially purified ISOM obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of ISOM. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of ISOM either by directly interacting with ISOM or by acting on components of the biological pathway in which ISOM participates.

An "allelic variant" is an alternative form of the gene encoding ISOM. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding ISOM include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as ISOM or a polypeptide with at least one functional characteristic of ISOM. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding ISOM, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding ISOM. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent ISOM. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the

residues, as long as the biological or immunological activity of ISOM is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of ISOM. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of ISOM either by directly interacting with ISOM or by acting on components of the biological pathway in which ISOM participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind ISOM polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA;

RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic ISOM, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding ISOM or fragments of ISOM may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (PE Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows

amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

A "fragment" is a unique portion of ISOM or the polynucleotide encoding ISOM which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment

used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50% of a polypeptide) as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:9-16 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:9-16, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:9-16 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:9-16 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:9-16 and the region of SEQ ID NO:9-16 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-8 is encoded by a fragment of SEQ ID NO:9-16. A fragment of SEQ ID NO:1-8 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-8. For example, a fragment of SEQ ID NO:1-8 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-8. The precise length of a fragment of SEQ ID NO:1-8 and the region of SEQ ID NO:1-8 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full-length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full-length" polynucleotide sequence encodes a "full-length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of

molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters ar set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode

similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (Apr-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for

chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of ISOM which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of ISOM which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of ISOM. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of ISOM.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of

amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an ISOM may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of ISOM.

"Probe" refers to nucleic acid sequences encoding ISOM, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of

Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and

other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequenc of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding ISOM, or fragments thereof, or ISOM itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of

replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants, and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook, J. et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-

1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human isomerases (ISOM), the polynucleotides encoding ISOM, and the use of these compositions for the diagnosis, treatment, or prevention of immune and cell proliferation disorders including cancer.

Table 1 lists the Incyte clones used to assemble full length nucleotide sequences encoding ISOM. Columns 1 and 2 show the sequence identification numbers (SEQ ID NOs) of the polypeptide and nucleotide sequences, respectively. Column 3 shows the clone IDs of the Incyte clones in which nucleic acids encoding each ISOM were identified, and column 4 shows the cDNA libraries from which these clones were isolated. Column 5 shows Incyte clones and their corresponding cDNA libraries. Clones for which cDNA libraries are not indicated were derived from pooled cDNA libraries. The Incyte clones in column 5 were used to assemble the consensus nucleotide sequence of each ISOM and are useful as fragments in hybridization technologies.

The columns of Table 2 show various properties of each of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each polypeptide; column 3 shows potential phosphorylation sites; column 4 shows potential glycosylation sites; column 5 shows the amino acid residues comprising signature sequences and motifs; column 6 shows homologous sequences as identified by BLAST analysis; and column 7 shows analytical methods and in some cases, searchable databases to which the analytical methods were applied. The methods of column 7 were used to characterize each polypeptide through sequence homology and protein motifs.

The columns of Table 3 show the tissue-specificity and diseases, disorders, or conditions associated with nucleotide sequences encoding ISOM. The first column of Table 3 lists the nucleotide SEQ ID NOs. Column 2 lists fragments of the nucleotide sequences of column 1. These fragments are useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:9-16 and to distinguish between SEQ ID NO:9-16 and related polynucleotide sequences. The polypeptides encoded by these fragments are useful, for example, as immunogenic peptides. Column 3 lists tissue categories which express ISOM as a fraction of total tissues expressing ISOM. Column 4 lists diseases, disorders, or conditions associated with those tissues expressing ISOM as a fraction of total tissues expressing ISOM. Column 5 lists the vectors used to subclone each cDNA library.

The columns of Table 4 show descriptions of the tissues used to construct the cDNA libraries from which cDNA clones encoding ISOM were isolated. Column 1 references the nucleotide SEQ ID NOs, column 2 shows the cDNA libraries from which these clones were isolated, and column 3 shows

the tissue origins and other descriptive information relevant to the cDNA libraries in column 2.

SEQ ID NO:13 maps to chromosome 2 within the interval from 188.2 to 201.7. SEQ ID NO:16 maps to chromosome 16 within the interval from 19.70 to 33.30. centiMorgans. The interval on chromosome 16 from 19.70 to 21.80 centiMorgans also contains the gene and ESTs encoding the protein stannin. The presence of stannin in a cell renders it sensitive to the effects of the drug trimethyltin (TMT) which induces neuronal damage in the brain of humans (Krady, J.K. et al. (1990) Brain Res. Molec. Brain Res. 7:287-297 and Toggas, S.M. et al. (1992) Molec. Pharm. 42:44-56). Stannin has also been shown to expressed in atherosclerotic lesions when activated by tumor necrosis factor-alpha (Online Mendelian Inheritance in Man (OMIM) *603032 Stannin; SNN; Horrevoets, A.J. et al. (1999) Blood 93:3418-3431). The interval on chromosome 16 from 27.00 to 34.60 centiMorgans also contains the gene encoding multidrug resistance-associated protein (MRP) as mapped in a small cell lung carcinoma cell line NCI-H69. MRP has sequence similarity to an ATP-binding cassette superfamily of transport systems which include the genes encoding the transmembrane transport protein P-glycoprotein (MDR1) and the cystic fibrosis transmembrane conductance regulator (CFTR) (OMIM *158343 Multidrug Resistance-Associated Protein 1; MRP1; Cole, S.P.C. et al. (1992) Science 258:1650-1654).

The invention also encompasses ISOM variants. A preferred ISOM variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the ISOM amino acid sequence, and which contains at least one functional or structural characteristic of ISOM.

The invention also encompasses polynucleotides which encode ISOM. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:9-16, which encodes ISOM. The polynucleotide sequences of SEQ ID NO:9-16, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding ISOM. In particular, such a variant polynucleotide sequence will have at least about 80%, or alternatively at least about 90%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding ISOM. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:9-16 which has at least about 80%, or alternatively at least about 90%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:9-16. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of ISOM.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding ISOM, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring ISOM, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode ISOM and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring ISOM under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding ISOM or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding ISOM and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode ISOM and ISOM derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding ISOM or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:9-16 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (PE Biosystems, Foster City CA), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably,

sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (PE Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (PE Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding ISOM may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, PE Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode ISOM may be cloned in recombinant DNA molecules that direct expression of ISOM, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express ISOM.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter ISOM-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of ISOM, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different speci s, thereby

maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding ISOM may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, ISOM itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (PE Biosystems). Additionally, the amino acid sequence of ISOM, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active ISOM, the nucleotide sequences encoding ISOM or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding ISOM. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding ISOM. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding ISOM and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding ISOM and appropriate transcriptional and translational control

elements. These meth ds include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding ISOM. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding ISOM. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding ISOM can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding ISOM into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of ISOM are needed, e.g. for the production of

antibodies, vectors which direct high level expression of ISOM may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of ISOM. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, supra; and Scorer, supra.)

Plant systems may also be used for expression of ISOM. Transcription of sequences encoding ISOM may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, supra; Broglie, supra; and Winter, supra.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding ISOM may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses ISOM in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of ISOM in cell lines is preferred. For example, sequences encoding ISOM can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance

to a selective agent, and its presenc allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ and *apr*⁻ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding ISOM is inserted within a marker gene sequence, transformed cells containing sequences encoding ISOM can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding ISOM under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding ISOM and that express ISOM may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of ISOM using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on ISOM is preferred, but a

competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immun logy, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding ISOM include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding ISOM, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding ISOM may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode ISOM may be designed to contain signal sequences which direct secretion of ISOM through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding ISOM may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric ISOM protein

containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of ISOM activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the ISOM encoding sequence and the heterologous protein sequence, so that ISOM may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled ISOM may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

ISOM of the present invention or fragments thereof may be used to screen for compounds that specifically bind to ISOM. At least one and up to a plurality of test compounds may be screened for specific binding to ISOM. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of ISOM, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which ISOM binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express ISOM, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing ISOM or cell membrane fractions which contain ISOM are then contacted with a test compound and binding, stimulation, or inhibition of activity of either ISOM or the compound is analyzed.

An assay may simply test binding of a test compound to th polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with ISOM, either in solution or affixed to a solid support, and detecting the binding of ISOM to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

ISOM of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of ISOM. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for ISOM activity, wherein ISOM is combined with at least one test compound, and the activity of ISOM in the presence of a test compound is compared with the activity of ISOM in the absence of the test compound. A change in the activity of ISOM in the presence of the test compound is indicative of a compound that modulates the activity of ISOM. Alternatively, a test compound is combined with an in vitro or cell-free system comprising ISOM under conditions suitable for ISOM activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of ISOM may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding ISOM or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding ISOM may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding ISOM can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding ISOM is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress ISOM, e.g., by secreting ISOM in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of ISOM and isomerases. In addition, the expression of ISOM is closely associated with reproductive, hematopoietic/immune, and gastrointestinal tissues. Therefore, ISOM appears to play a role in immune and cell proliferation disorders including cancer. In the treatment of disorders associated with increased ISOM expression or activity, it is desirable to decrease the expression or activity of ISOM. In the treatment of disorders associated with decreased ISOM expression or activity, it is desirable to increase the expression or activity of ISOM.

Therefore, in one embodiment, ISOM or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of ISOM. Examples of such disorders include, but are not limited to, an immune disorder such as inflammation, actinic keratosis, acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, arteriosclerosis, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, bursitis, cholecystitis, cirrhosis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, paroxysmal nocturnal hemoglobinuria, hepatitis, hypereosinophilia, irritable bowel syndrome, episodic lymphopenia with lymphocytotoxins, mixed connective tissue disease (MCTD), multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, myelofibrosis, osteoarthritis, osteoporosis, pancreatitis, polycythemia vera, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic

lupus eryth matosus, systemic sclerosis, primary thrombocythemia, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, trauma, and hematopoietic cancer including lymphoma, leukemia, and myeloma; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing ISOM or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of ISOM including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified ISOM in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of ISOM including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of ISOM may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of ISOM including, but not limited to, those listed above.

In a further embodiment, an antagonist of ISOM may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of ISOM. Examples of such disorders include, but are not limited to, those immune and cell proliferation disorders including cancer described above. In one aspect, an antibody which specifically binds ISOM may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express ISOM.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding ISOM may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of ISOM including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic

efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of ISOM may be produced using methods which are generally known in the art. In particular, purified ISOM may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind ISOM. Antibodies to ISOM may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with ISOM or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to ISOM have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of ISOM amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to ISOM may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce ISOM-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g.,

Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for ISOM may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between ISOM and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering ISOM epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for ISOM. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of ISOM-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple ISOM epitopes, represents the average affinity, or avidity, of the antibodies for ISOM. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular ISOM epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the ISOM-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of ISOM, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For

xample, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of ISOM-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al., supra.)

In another embodiment of the invention, the polynucleotides encoding ISOM, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding ISOM. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding ISOM. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding ISOM may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii)

express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) xpress a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in ISOM expression or regulation causes disease, the expression of ISOM from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in ISOM are treated by constructing mammalian expression vectors encoding ISOM and introducing these vectors by mechanical means into ISOM-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of ISOM include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). ISOM may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding ISOM from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al.

(1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to ISOM expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding ISOM under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding ISOM to cells which have one or more genetic abnormalities with respect to the expression of ISOM. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544; and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding ISOM to target cells which have one or more genetic abnormalities with

respect to the expression of ISOM. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing ISOM to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res.169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding ISOM to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full-length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for ISOM into the alphavirus genome in place of the capsid-coding region results in the production of a large number of ISOM-coding RNAs and the synthesis of high levels of ISOM in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of ISOM into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the

art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding ISOM.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding ISOM. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine,

queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding ISOM. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased ISOM expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding ISOM may be therapeutically useful, and in the treament of disorders associated with decreased ISOM expression or activity, a compound which specifically promotes expression of the polynucleotide encoding ISOM may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding ISOM is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding ISOM are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding ISOM. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res.

Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of ISOM, antibodies to ISOM, and mimetics, agonists, antagonists, or inhibitors of ISOM.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising ISOM or fragments ther of. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion intracellular delivery of the macromolecule. Alternatively, ISOM or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example ISOM or fragments thereof, antibodies of ISOM, and agonists, antagonists or inhibitors of ISOM, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their

inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind ISOM may be used for the diagnosis of disorders characterized by expression of ISOM, or in assays to monitor patients being treated with ISOM or agonists, antagonists, or inhibitors of ISOM. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for ISOM include methods which utilize the antibody and a label to detect ISOM in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring ISOM, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of ISOM expression. Normal or standard values for ISOM expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibody to ISOM under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of ISOM expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding ISOM may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of ISOM may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of ISOM, and to monitor regulation of ISOM levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding ISOM or closely related molecules may be used to identify nucleic acid sequences which encode ISOM. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding ISOM, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the ISOM encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:9-16 or from

genomic sequences including promoters, enhancers, and introns of the ISOM gene.

Means for producing specific hybridization probes for DNAs encoding ISOM include the cloning of polynucleotide sequences encoding ISOM or ISOM derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding ISOM may be used for the diagnosis of disorders associated with expression of ISOM. Examples of such disorders include, but are not limited to, an immune disorder such as inflammation, actinic keratosis, acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, arteriosclerosis, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, bursitis, cholecystitis, cirrhosis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, paroxysmal nocturnal hemoglobinuria, hepatitis, hypereosinophilia, irritable bowel syndrome, episodic lymphopenia with lymphocytotoxins, mixed connective tissue disease (MCTD), multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, myelofibrosis, osteoarthritis, osteoporosis, pancreatitis, polycythemia vera, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, primary thrombocythemia, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, trauma, and hematopoietic cancer including lymphoma, leukemia, and myeloma; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding ISOM may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered ISOM expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding ISOM may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding ISOM may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding ISOM in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of ISOM, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding ISOM, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding ISOM may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding ISOM, or a fragment of a polynucleotide complementary to the polynucleotide encoding ISOM, and will be employed under optimized conditions for identification of a specific gene or

condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding ISOM may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding ISOM are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of ISOM include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described in Seilhamer, J.J. et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, incorporated herein by reference. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and

monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, antibodies specific for ISOM, or ISOM or fragments thereof may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful f r

comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may als be generated using antibodies specific for ISOM to quantify the levels of ISOM expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to th sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-

2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference. .

In another embodiment of the invention, nucleic acid sequences encoding ISOM may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, e.g., Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding ISOM on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, ISOM, its catalytic or immunogenic fragments, or

oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affix d to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between ISOM and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with ISOM, or fragments thereof, and washed. Bound ISOM is then detected by methods well known in the art. Purified ISOM can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding ISOM specifically compete with a test compound for binding ISOM. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with ISOM.

In additional embodiments, the nucleotide sequences which encode ISOM may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, in particular U.S. Ser. No.60/149,388, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), pcDNA2.1 plasmid (Invitrogen, Carlsbad CA), or pINCY plasmid (Incyte Genomics, Palo Alto CA). Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (PE Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (PE Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VI.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 5 summarizes the tools, programs, and algorithms used and provides applicable descriptions, references, and threshold parameters. The first column of Table 5 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments were generated using the default parameters specified by the clustal algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programing, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and PFAM to acquire annotation using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled

into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:9-16. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

IV. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced

either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding ISOM occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation, trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in Table 3.

V. Chromosomal Mapping of ISOM Encoding Polynucleotides

The cDNA sequences which were used to assemble SEQ ID NO:9-16 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:9-16 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 5). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

The genetic map locations of SEQ ID NO:13 and SEQ ID NO:16 are described in The Invention as ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination). The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Diseases associated with the public and Incyte sequences located within the indicated intervals are also reported in the Invention where applicable.

VI. Extension of ISOM Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:9-16 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5'extension of the known fragment, and the other

primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68 °C to about 72 °C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and β-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94 °C, 3 min; Step 2: 94 °C, 15 sec; Step 3: 60 °C, 1 min; Step 4: 68 °C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68 °C, 5 min; Step 7: storage at 4 °C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94 °C, 3 min; Step 2: 94 °C, 15 sec; Step 3: 57 °C, 1 min; Step 4: 68 °C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68 °C, 5 min; Step 7: storage at 4 °C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37 °C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; St p 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems).

In like manner, the polynucleotide sequences of SEQ ID NO:9-16 are used to obtain 5' regulatory sequences using the procedure above, along with oligonucleotides designed for such extension, and an appropriate genomic library.

VII. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:9-16 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

VIII. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999),

supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/µl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/µl RNase inhibitor, 500 µM dATP, 500 µM dGTP, 500 µM dTTP, 40 µM dCTP, 40 µM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37 °C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85 °C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 µl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60 °C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65 °C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60 °C. The arrays are washed for 10 min at 45 °C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45 °C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines

at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

IX. Complementary Polynucleotides

Sequences complementary to the ISOM-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring ISOM. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with

smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of ISOM. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the ISOM-encoding transcript.

X. Expression of ISOM

Expression and purification of ISOM is achieved using bacterial or virus-based expression systems. For expression of ISOM in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express ISOM upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of ISOM in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding ISOM by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, ISOM is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from ISOM at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified ISOM obtained by these methods can be used directly in the assays shown in Examples XI and XV.

XI. Demonstrati n of ISOM Activity

ISOM activity is demonstrated through a variety of specific enzyme assays, some of which are outlined below.

Peptidyl Prolyl *cis-trans* Isomerase Activity

ISOM-6 peptidyl prolyl *cis-trans* isomerase activity can be assayed as described (Rahfeld, J.U. et al. (1994) FEBS Lett. 352:180-184). The assay is performed at 10°C in 35 mM HEPES buffer, pH 7.8, containing chymotrypsin (0.5 mg/ml) and ISOM-6 at a variety of concentrations. In this assay, the substrate is a peptide containing four hydrophobic residues. The peptide contains a succinate group at the N-terminus and a nitroanilide group at the C-terminus. The substrate is in equilibrium with respect to the prolyl bond, with 80-95% in *trans* and 5-20% in *cis* conformation. An aliquot (2 µl) of the substrate dissolved in dimethyl sulfoxide (10 mg/ml) is added to the reaction mixture described above. Only the *cis* isomer of the substrate is a substrate for cleavage by chymotrypsin. Thus, as the substrate is isomerized by ISOM-6, the product is cleaved by chymotrypsin to produce 4-nitroanilide, which is detected by its absorbance at 390 nm. 4-Nitroanilide appears in a time-dependent and an ISOM-6 concentration-dependent manner.

Alternatively, peptidyl prolyl *cis-trans* isomerase activity of ISOM-6 can be assayed using a chromogenic peptide in a coupled assay with chymotrypsin (Fischer, G. et al. (1984) Biomed. Biochim. Acta 43:1101-1111).

Thioredoxin Activity

ISOM thioredoxin activity is assayed as described (Luthman, M. (1982) Biochemistry 21:6628-6633). Thioredoxins catalyze the formation of disulfide bonds and regulate the redox environment in cells to enable the necessary thiol:disulfide exchanges. One way to measure the thiol:disulfide exchange is by measuring the reduction of insulin in a mixture containing 0.1M potassium phosphate, pH 7.0, 2 mM EDTA, 0.16 µM insulin, 0.33 mM DTT, and 0.48 mM NADPH. Different concentrations of ISOM are added to the mixture, and the reaction rate is followed by monitoring the oxidation of NADPH at 340 nM.

Transferase Activity

ISOM transferase activity is measured through a methyl transferase assay in which the transfer of radiolabeled methyl groups between a donor substrate and an acceptor substrate is measured (Bokar, J.A. et al. (1994) J. Biol. Chem. 269:17697-17704). Reaction mixtures (50 µl final volume) contain 15 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 10 mM dithiothreitol, 3% polyvinylalcohol, donor substrate (1.5 µCi [*methyl*-^{3}H]AdoMet (0.375 µM AdoMet, DuPont-NEN), 0.6 µg ISOM, and acceptor substrate (0.4 µg [^{35}S]RNA or 6-mercaptopurine (6-MP) to 1 mM final concentration). Reaction mixtures are incubated at 30°C for 30 minutes, then 65°C for 5 minutes. The products are separated by chromatography or electrophoresis and the level of methyl transferase activity is

determined by quantification of *methyl*-^{3}H-RNA or *methyl*-^{3}H-6-MP recovery.

XII. Functional Assays

ISOM function is assessed by expressing the sequences encoding ISOM at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT plasmid (Life Technologies) and pCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of ISOM on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding ISOM and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding ISOM and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIII. Production of ISOM Specific Antibodies

ISOM substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to pr duce antibodies using standard protocols.

Alternatively, the ISOM amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (PE Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-ISOM activity by, for example, binding the peptide or ISOM to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XIV. Purification of Naturally Occurring ISOM Using Specific Antibodies

Naturally occurring or recombinant ISOM is substantially purified by immunoaffinity chromatography using antibodies specific for ISOM. An immunoaffinity column is constructed by covalently coupling anti-ISOM antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing ISOM are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of ISOM (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/ISOM binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and ISOM is collected.

XV. Identification of Molecules Which Interact with ISOM

ISOM, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled ISOM, washed, and any wells with labeled ISOM complex are assayed. Data obtained using different concentrations of ISOM are used to calculate values for the number, affinity, and association of ISOM with the candidate molecules.

Alternatively, molecules interacting with ISOM are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989, Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

ISOM may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions

between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	9	011886	THP1PLB01	011886H1 (THP1PLB01), 036126X29R1 (HUVENOB01), 036126X33R1 (HUVENOB01), 990094T1 (COLNNOT11), 1998624H1 (BRSTTUT03), 3355930H1 (PROSTUT16), SAIA01790F1, SAIA02403F1
2	10	1863189	PROSNOT19	764845R6 (LUNGNOT04), 1863189F6 (PROSNOT19), 1863189H1 (PROSNOT19), 1863189T6 (PROSNOT19), 1975669F6 (PANCTUT02), 2780516F6 (OVARTUT03), 2937335H1 (THYMFET02), 3890112H1 (BRSTTUT16), 4823918H1 (BLADDIT01), 4946302H1 (SINTNOT25)
3	11	2088868	PANCNOT04	222449R1 (PANCNOT01), 2088868H1 (PANCNOT04), 2088868T6 (PANCNOT04), 2773734F6 (PANCNOT15)
4	12	2481256	SMCANOT01	550022T6 (BEPINOT01), 571246H1 (MMLR3DT01), 1867467F6 (SKINBIT01), 2481256H1 (SMCANOT01), 4976719H1 (HELATXT03)
5	13	2505257	CONUTUT01	642581R1 (BRSTTUT02), 660543T6 (BRAINOT03), 1255731H1 (MENITUT03), 1396178T6 (THYRNOT03), 1440452F1 (THYRNOT03), 1440452R1 (THYRNOT03), 1466643F6 (PANCTUT02), 2068909F6 (ISLTNOT01), 2134626T6 (ENDCNOT01), 2231607H1 (PROSNOT16), 2505257F6 (CONUTUT01), 2505257H1 (CONUTUT01), 3586111H1 (293TF4T01), SAEA01132F1, SAEA10039P1, SAEA03368F1
6	14	3325534	PTHYNOT03	1822950F6 (GBLATUT01), 3325534H1 (PTHYNOT03), SAEA00032R1, SAEA01058R1, SBMA03334F1, SAEA00977F1

Table 1 (cont.)

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
7	15	3817050	TONSNOT03	3817050H1 (TONSNOT03), SAEA00998F1, SAEA01958R1, SAEA00601R1, SAEA01633F1
8	16	5324378	FIBPFEN06	809609T1 (LUNGNOT04), 1359567F1 (LUNGNOT12), 1443604R1 (THYRNOT03), 1597902F6 (BLADNOT03), 1684060F6 (PROSNOT15), 1878192F6 (LEUKNOT03), 2071801F6 (ISLTNOT01), 2192619F6 (THYRTUT03), 4747073H2 (SMCRUNT01), 5324378H1 (FIBPFEN06)

Table 2

Poly-p ptide Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation sites	Signature Sequence	Homologous Sequences	Analytical Methods & Databases
1	542	T30 S86 S110 S139 S201 T381 S463 T465 S498 S133 S207 S334 Y184 Y227	N331	Phosphoglucomutase and phosphomannomutase signature N57-K74 N269-Y284 N-acetylglucosamine-phosphate isomerase N96-E339 T349-V523 DNA-damage-repair/ toleration protein, DRT101 precursor L338-G532	Phosphomannomutase phosphoserine (g908894) Schizosaccharomyces pombe	BLAST-GenBank HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO MOTIFS
2	311	S202 S61 S76 T145 S147 S230	N57	Ribose 5-phosphate isomerase M75-F310	Ribose 5-phosphate isomerase (g836674) Mus musculus	BLAST-GenBank BLAST-PRODOM BLAST-DOMO MOTIFS
3	273	T91 T227 S254 T268	N100	Signal peptide M1-A25 Protein isomerase precursor signal A49-E260 Protein disulfide-isomerase K29-E265	Protein disulfide isomerase (E.C.5.3.4.1) (g163497) Bos taurus	BLAST-GenBank HMMER BLAST-PRODOM BLAST-DOMO MOTIFS

Table 2 (cont.)

Poly-peptide Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation sites	Signature Sequence	Homologous Sequences	Analytical Methods & Databases
4	228	S29 S74 T123 S56	N46	Ribulose-phosphate 3-epimerase family K6-L226	Ribulose-5-phosphate-epimerase (g2894532) Homo sapiens	BLAST-GenBank BLIMPS-BLOCKS BLAST-PRODOM BLAST-DOMO MOTIFS
5	793	S494 T32 S47 T167 T236 T357 S405 T509 T565 T679 T571 S641 S659 S681 T745 S776 T781	N530	Thioredoxins D128-R234 N452-P638 P669-E780 Isomerase L134-V233	Putative protein disulfide isomerase precursor (g2702281) Arapidopsis thaliana	BLAST-GenBank HMMER-PFAM BLIMPS-BLOCKS PROFILECAN BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO MOTIFS
6	492	T93 T389 T8 S76 T111 T142 S204 T251 T252 T332 T361 T407 S425 T59 S188 T381 S382 S432 S456 S471 S479 Y283 Y363 Y453 Y491	N99 N140 N157 N440	RNA recognition motif L242-V313 Cyclophilin-type peptidyl-prolyl cis-trans isomerase signature M1-F166 Multigene family cyclophilin protein L9-D145	Similarity to Brugia peptidylprolyl isomerase (g3420982) Caenoreabacterium elegans	BLAST-GenBank HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO MOTIFS

Table 2 (cont.)

Poly-peptide Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation sites	Signature Sequence	Homologous Sequences	Analytical Methods & Databases
7	160	T149	N9 N105	Cyclophilin-type peptidyl-prolyl cis-trans isomerase signature T5-L160 Multigene family cyclophilin protein V2-K151	Peptidylprolyl isomerase (g30168) Homo sapiens	BLAST-GenBank HMMER-PFAM BLIMPS-BLOCKS PROFILESCAN BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO MOTIFS
8	744	S250 S321 T458 S508 S587 S47 T55 T62 S359 T362 S535 S616 S735	N13 N85 N96 N233 N354 N363 N425 N728 E741	Thioredoxin family signature Q448-D558		BLAST-GenBank HMMER-PFAM PROFILESCAN BLIMPS-PRINTS BLAST-DOMO MOTIFS

Table 3

Nucleotide Seq ID NO:	Selected Fragments(s)	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
9	432-476 834-878	Reproductive (0.289) Cardiovascular (0.158) Gastrointestinal (0.132) Musculoskeletal (0.132)	Cancer (0.474) Inflammation/Trauma (0.289) Cell Proliferation (0.158)	PBLUESCRIPT
10	208-252	Hematopoietic/Immune (0.265) Reproductive (0.245) Gastrointestinal (0.163)	Cancer (0.449) Cell Proliferation (0.265) Inflammation/Trauma (0.347)	pINCY
11	267-311	Gastrointestinal (0.333) Hematopoietic/Immune (0.200) Urologic (0.200)	Cancer (0.333) Inflammation/Trauma (0.600) Cell Proliferation (0.267)	PSPORT1
12	645-689 900-944	Cardiovascular (0.200) Hematopoietic/Immune (0.171) Reproductive (0.171)	Cancer (0.429) Inflammation (0.314) Cell Proliferation (0.143)	pINCY

Table 3 (cont.)

Nucleotide Seq ID NO:	Selected Fragments(s)	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
13	562-606 1447-1491 2020-2064	Reproductive (0.259) Gastrointestinal (0.207) Cardiovascular (0.164)	Cancer (0.517) Inflammation/Trauma (0.302) Cell Proliferation (0.190)	pINCY
14	548-592	Gastrointestinal (0.203) Reproductive (0.203) Nervous (0.153)	Cancer (0.458) Inflammation/Trauma (0.373) Cell Proliferation (0.203)	pINCY
15	22-66 520-564	Reproductive (0.279) Nervous (0.131) Hematopoietic/Immune (0.115)	Cancer (0.377) Inflammation/Trauma (0.393) Cell Proliferation (0.197)	pINCY
16	344-388 995-1039 1913-1957 2435-2479	Reproductive (0.274) Hematopoietic/Immune (0.155) Nervous (0.131)	Cancer (0.476) Inflammation/Trauma (0.333) Cell Proliferation (0.167)	pINCY

Table 4

Nucleotide SEQ ID NO:	Library	Library Comment
9	THP1PLB01	Library was constructed using RNA isolated from THP-1 cells cultured for 48 hours with 100 ng/ml phorbol ester (PMA), followed by a 4-hour culture in media containing 1 ug/ml LPS. THP-1 (ATCC TIB 202) is a human promonocyte line derived from the peripheral blood of a 1-year-old male with acute monocytic leukemia.
10	PROSNOT19	Library was constructed using RNA isolated from diseased prostate tissue removed from a 59-year-old Caucasian male during a radical prostatectomy with regional lymph node excision. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma (Gleason grade 3+3). The patient presented with elevated prostate-specific antigen (PSA). Patient history included colon diverticuli, asbestosis, and thrombophlebitis. Previous surgeries included a partial colectomy. Family history included benign hypertension, multiple myeloma, hyperlipidemia and rheumatoid arthritis.
11	PANCNOT04	Library was constructed using RNA isolated from the pancreatic tissue of a 5-year-old Caucasian male who died in a motor vehicle accident.
12	SMCANOT01	Library was constructed using RNA isolated from an aortic smooth muscle cell line derived from the explanted heart of a male during a heart transplant.

Table 4 (cont.)

Nucleotide SEQ ID NO:	Library	Library Comment
13	CONUTUT01	Library was constructed using RNA isolated from sigmoid mesentery tumor tissue obtained from a 61-year-old female during a total abdominal hysterectomy and bilateral salpingo-oophorectomy with regional lymph node excision. Pathology indicated a metastatic grade 4 malignant mixed mullerian tumor present in the sigmoid mesentery at two sites.
14	PTHYNOT03	Library was constructed using RNA isolated from the left parathyroid tissue of a 69-year-old Caucasian female during a partial parathyroidectomy. Pathology indicated hyperplasia. The patient presented with primary hyperparathyroidism.
15	TONSNOT03	Library was constructed using RNA isolated from diseased left tonsil tissue removed from a 6-year-old Caucasian male during adenotonsillectomy. Pathology indicated reactiv lymphoid hyperplasia, bilaterally. Family history included benign hypertension, myocardial infarction, and atherosclerotic coronary artery disease.
16	FIBPFEN06	Library was constructed using RNA isolated from the prostate stroma removed from a male fetus, who died after 26 weeks gestation.

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	PE Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	PE Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	PE Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and, if applicable, Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits for PFAM hits, depending on individual protein families

Table 5 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

a) an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8,

b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8,

c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, and

d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method for producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence op rably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16,

b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16,

c) a polynucleotide sequence complementary to a),

d) a polynucleotide sequence complementary to b), and

e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequenc of a polynucleotide of claim 11, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A composition of claim 16, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8.

18. A method for treating a disease or condition associated with decreased expression of functional ISOM, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

21. A method for treating a disease or condition associated with decreased expression of functional ISOM, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional ISOM, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, and

b) detecting altered expression of the target polynucleotide.

28. A method for assessing toxicity of a test compound, said method comprising:

a) treating a biological sample containing nucleic acids with the test compound;

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;

c) quantifying the amount of hybridization complex; and

d) comparing the amount of hybridization complex in th treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

SEQUENCE LISTING

<110> INCYTE GENOMICS, INC.
BANDMAN, Olga
LU, Dyung Aina M.
YUE, Henry
TRAN, Bao
HILLMAN, Jennifer L.
BAUGHN, Mariah R.
LAL, Preeti
TANG, Y. Tom

<120> ISOMERASE PROTEINS

<130> PF-0730 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/149,388
<151> 1999-08-17

<160> 16

<170> PERL Program

<210> 1
<211> 542
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 011886CD1

<400> 1
Met Asp Leu Gly Ala Ile Thr Lys Tyr Ser Ala Leu His Ala Lys
1 5 10 15
Pro Asn Gly Leu Ile Leu Gln Tyr Gly Thr Ala Gly Phe Arg Thr
20 25 30
Lys Ala Glu His Leu Asp His Val Met Phe Arg Met Gly Leu Leu
35 40 45
Ala Val Leu Arg Ser Lys Gln Thr Lys Ser Thr Ile Gly Val Met
50 55 60
Val Thr Ala Ser His Asn Pro Glu Glu Asp Asn Gly Val Lys Leu
65 70 75
Val Asp Pro Leu Gly Glu Met Leu Ala Pro Ser Trp Glu Glu His
80 85 90
Ala Thr Cys Leu Ala Asn Ala Glu Glu Gln Asp Met Gln Arg Val
95 100 105
Leu Ile Asp Ile Ser Glu Lys Glu Ala Val Asn Leu Gln Gln Asp
110 115 120
Ala Phe Val Val Ile Gly Arg Asp Thr Arg Pro Ser Ser Glu Lys
125 130 135
Leu Ser Gln Ser Val Ile Asp Gly Val Thr Val Leu Gly Gly Gln
140 145 150
Phe His Asp Tyr Gly Leu Leu Thr Thr Pro Gln Leu His Tyr Met
155 160 165
Val Tyr Cys Arg Asn Thr Gly Gly Arg Tyr Gly Lys Ala Thr Ile
170 175 180
Glu Gly Tyr Tyr Gln Lys Leu Ser Lys Ala Phe Val Glu Leu Thr

```
                185                 190                 195
Lys Gln Ala Ser Cys Ser Gly Asp Glu Tyr Arg Ser Leu Lys Val
                200                 205                 210
Asp Cys Ala Asn Gly Ile Gly Ala Leu Lys Leu Arg Glu Met Glu
                215                 220                 225
His Tyr Phe Ser Gln Gly Leu Ser Val Gln Leu Phe Asn Asp Gly
                230                 235                 240
Ser Lys Gly Lys Leu Asn His Leu Cys Gly Ala Asp Phe Val Lys
                245                 250                 255
Ser His Gln Lys Pro Pro Gln Gly Met Glu Ile Lys Ser Asn Glu
                260                 265                 270
Arg Cys Cys Ser Phe Asp Gly Asp Ala Asp Arg Ile Val Tyr Tyr
                275                 280                 285
Tyr His Asp Ala Asp Gly His Phe His Leu Ile Asp Gly Asp Lys
                290                 295                 300
Ile Ala Thr Leu Ile Ser Ser Phe Leu Lys Glu Leu Leu Val Glu
                305                 310                 315
Ile Gly Glu Ser Leu Asn Ile Gly Val Val Gln Thr Ala Tyr Ala
                320                 325                 330
Asn Gly Ser Ser Thr Arg Tyr Leu Glu Glu Val Met Lys Val Pro
                335                 340                 345
Val Tyr Cys Thr Lys Thr Gly Val Lys His Leu His His Lys Ala
                350                 355                 360
Gln Glu Phe Asp Ile Gly Val Tyr Phe Glu Ala Asn Gly His Gly
                365                 370                 375
Thr Ala Leu Phe Ser Thr Ala Val Glu Met Lys Ile Lys Gln Ser
                380                 385                 390
Ala Glu Gln Leu Glu Asp Lys Lys Arg Lys Ala Ala Lys Met Leu
                395                 400                 405
Glu Asn Ile Ile Asp Leu Phe Asn Gln Ala Ala Gly Asp Ala Ile
                410                 415                 420
Ser Asp Met Leu Val Ile Glu Ala Ile Leu Ala Leu Lys Gly Leu
                425                 430                 435
Thr Val Gln Gln Trp Asp Ala Leu Tyr Thr Asp Leu Pro Asn Arg
                440                 445                 450
Gln Leu Lys Val Gln Val Ala Asp Arg Arg Val Ile Ser Thr Thr
                455                 460                 465
Asp Ala Glu Arg Gln Ala Val Thr Pro Pro Gly Leu Gln Glu Ala
                470                 475                 480
Ile Asn Asp Leu Val Lys Lys Tyr Lys Leu Ser Arg Ala Phe Val
                485                 490                 495
Arg Pro Ser Gly Thr Glu Asp Val Val Arg Val Tyr Ala Glu Ala
                500                 505                 510
Asp Ser Gln Glu Ser Ala Asp His Leu Ala His Glu Val Ser Leu
                515                 520                 525
Ala Val Phe Gln Leu Ala Gly Gly Ile Gly Glu Arg Pro Gln Pro
                530                 535                 540
Gly Phe
```

<210> 2
<211> 311
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1863189CD1

<400> 2
Met Gln Arg Pro Gly Pro Phe Ser Thr Leu Tyr Gly Arg Val Leu

```
  1               5                  10                  15
Ala Pro Leu Pro Gly Arg Ala Gly Gly Ala Ala Ser Gly Gly Gly
                20                  25                  30
Gly Asn Ser Trp Asp Leu Pro Gly Ser His Val Arg Leu Pro Gly
                35                  40                  45
Arg Ala Gln Ser Gly Thr Arg Gly Gly Ala Gly Asn Thr Ser Thr
                50                  55                  60
Ser Cys Gly Asp Ser Asn Ser Ile Cys Pro Ala Pro Ser Thr Met
                65                  70                  75
Ser Lys Ala Glu Glu Ala Lys Lys Leu Ala Gly Arg Ala Ala Val
                80                  85                  90
Glu Asn His Val Arg Asn Asn Gln Val Leu Gly Ile Gly Ser Gly
                95                 100                 105
Ser Thr Ile Val His Ala Val Gln Arg Ile Ala Glu Arg Val Lys
               110                 115                 120
Gln Glu Asn Leu Asn Leu Val Cys Ile Pro Thr Ser Phe Gln Ala
               125                 130                 135
Arg Gln Leu Ile Leu Gln Tyr Gly Leu Thr Leu Ser Asp Leu Asp
               140                 145                 150
Arg His Pro Glu Ile Asp Leu Ala Ile Asp Gly Ala Asp Glu Val
               155                 160                 165
Asp Ala Asp Leu Asn Leu Ile Lys Gly Gly Gly Gly Cys Leu Thr
               170                 175                 180
Gln Glu Lys Ile Val Ala Gly Tyr Ala Ser Arg Phe Ile Val Ile
               185                 190                 195
Ala Asp Phe Arg Lys Asp Ser Lys Asn Leu Gly Asp Gln Trp His
               200                 205                 210
Lys Gly Ile Pro Ile Glu Val Ile Pro Met Ala Tyr Val Pro Val
               215                 220                 225
Ser Arg Ala Val Ser Gln Lys Phe Gly Gly Val Val Glu Leu Arg
               230                 235                 240
Met Ala Val Asn Lys Ala Gly Pro Val Val Thr Asp Asn Gly Asn
               245                 250                 255
Phe Ile Leu Asp Trp Lys Phe Asp Arg Val His Lys Trp Ser Glu
               260                 265                 270
Val Asn Thr Ala Ile Lys Met Ile Pro Gly Val Val Asp Thr Gly
               275                 280                 285
Leu Phe Ile Asn Met Ala Glu Arg Val Tyr Phe Gly Met Gln Asp
               290                 295                 300
Gly Ser Val Asn Met Arg Glu Lys Pro Phe Cys
               305                 310
```

<210> 3
<211> 273
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2088868CD1

<400> 3

```
Met Glu Ala Ala Pro Ser Arg Phe Met Phe Leu Leu Phe Leu Leu
  1               5                  10                  15
Thr Cys Glu Leu Ala Ala Glu Val Ala Ala Glu Val Glu Lys Ser
                20                  25                  30
Ser Asp Gly Pro Gly Ala Ala Gln Glu Pro Thr Trp Leu Thr Asp
                35                  40                  45
Val Pro Ala Ala Met Glu Phe Ile Ala Ala Thr Glu Val Ala Val
                50                  55                  60
```

```
Ile Gly Phe Phe Gln Asp Leu Glu Ile Pro Ala Val Pro Ile Leu
                65                  70                  75
His Ser Met Val Gln Lys Phe Pro Gly Val Ser Phe Gly Ile Ser
                80                  85                  90
Thr Asp Ser Glu Val Leu Thr His Tyr Asn Ile Thr Gly Asn Thr
                95                 100                 105
Ile Cys Leu Phe Arg Leu Val Asp Asn Glu Gln Leu Asn Leu Glu
               110                 115                 120
Asp Glu Asp Ile Glu Ser Ile Asp Ala Thr Lys Leu Ser Arg Phe
               125                 130                 135
Ile Glu Ile Asn Ser Leu His Met Val Thr Glu Tyr Asn Pro Val
               140                 145                 150
Thr Val Ile Gly Leu Phe Asn Ser Val Ile Gln Ile His Leu Leu
               155                 160                 165
Leu Ile Met Asn Lys Ala Ser Pro Glu Tyr Glu Glu Asn Met His
               170                 175                 180
Arg Tyr Gln Lys Ala Ala Lys Leu Phe Gln Gly Lys Ile Leu Phe
               185                 190                 195
Ile Leu Val Asp Ser Gly Met Lys Glu Asn Gly Lys Val Ile Ser
               200                 205                 210
Phe Phe Lys Leu Lys Glu Ser Gln Leu Pro Ala Leu Ala Ile Tyr
               215                 220                 225
Gln Thr Leu Asp Asp Glu Trp Asp Thr Leu Pro Thr Ala Glu Val
               230                 235                 240
Ser Val Glu His Val Gln Asn Phe Cys Asp Gly Phe Leu Ser Gly
               245                 250                 255
Lys Leu Leu Lys Glu Asn Arg Glu Ser Glu Gly Lys Thr Pro Lys
               260                 265                 270
Val Glu Leu
```

<210> 4
<211> 228
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2481256CD1

<400> 4

```
Met Ala Ser Gly Cys Lys Ile Gly Pro Ser Ile Leu Asn Ser Asp
  1               5                  10                  15
Leu Ala Asn Leu Gly Ala Glu Cys Leu Arg Met Leu Asp Ser Gly
               20                  25                  30
Ala Asp Tyr Leu His Leu Asp Val Met Asp Gly His Phe Val Pro
               35                  40                  45
Asn Ile Thr Phe Gly His Pro Val Val Glu Ser Leu Arg Lys Gln
               50                  55                  60
Leu Gly Gln Asp Pro Phe Phe Asp Met His Met Met Val Ser Lys
               65                  70                  75
Pro Glu Gln Trp Val Lys Pro Met Ala Val Ala Gly Ala Asn Gln
               80                  85                  90
Tyr Thr Phe His Leu Glu Ala Thr Glu Asn Pro Gly Ala Leu Ile
               95                 100                 105
Lys Asp Ile Arg Glu Asn Gly Met Lys Val Gly Leu Ala Ile Lys
              110                 115                 120
Pro Gly Thr Ser Val Glu Tyr Leu Ala Pro Trp Ala Asn Gln Ile
              125                 130                 135
Asp Met Ala Leu Val Met Thr Val Glu Pro Gly Phe Gly Gly Gln
              140                 145                 150
```

```
Lys Phe Met Glu Asp Met Met Pro Lys Val His Trp Leu Arg Thr
                155                 160                 165
Gln Phe Pro Ser Leu Asp Ile Glu Val Asp Gly Gly Val Gly Pro
                170                 175                 180
Asp Thr Val His Lys Cys Ala Glu Ala Gly Ala Asn Met Ile Val
                185                 190                 195
Ser Gly Ser Ala Ile Met Arg Ser Glu Asp Pro Arg Ser Val Ile
                200                 205                 210
Asn Leu Leu Arg Asn Val Cys Ser Glu Ala Ala Gln Lys Arg Ser
                215                 220                 225
Leu Asp Arg
```

<210> 5
<211> 793
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2505257CD1

<400> 5

```
Met Gly Val Trp Leu Asn Lys Asp Asp Asp Ile Arg Asp Leu Lys
  1               5                  10                  15
Arg Ile Ile Leu Cys Phe Leu Ile Val Tyr Met Ala Ile Leu Val
                20                  25                  30
Gly Thr Asp Gln Asp Phe Tyr Ser Leu Leu Gly Val Ser Lys Thr
                35                  40                  45
Ala Ser Ser Arg Glu Ile Arg Gln Ala Phe Lys Lys Leu Ala Leu
                50                  55                  60
Lys Leu His Pro Asp Lys Asn Pro Asn Asn Pro Asn Ala His Gly
                65                  70                  75
Asn Phe Leu Lys Ile Asn Arg Ala Tyr Glu Val Leu Lys Asp Glu
                80                  85                  90
Asp Leu Arg Lys Lys Tyr Asp Lys Tyr Gly Glu Lys Gly Leu Glu
                95                 100                 105
Asp Asn Gln Gly Gly Gln Tyr Glu Ser Trp Asn Tyr Tyr Arg Tyr
               110                 115                 120
Asp Phe Gly Ile Tyr Asp Asp Asp Pro Glu Ile Ile Thr Leu Glu
               125                 130                 135
Arg Arg Glu Phe Asp Ala Ala Val Asn Ser Gly Glu Leu Trp Phe
               140                 145                 150
Val Asn Phe Tyr Ser Pro Gly Cys Ser His Cys His Asp Leu Ala
               155                 160                 165
Pro Thr Trp Arg Asp Phe Ala Lys Glu Val Asp Gly Leu Leu Arg
               170                 175                 180
Ile Gly Ala Val Asn Cys Gly Asp Asp Arg Met Leu Cys Arg Met
               185                 190                 195
Lys Gly Val Asn Ser Tyr Pro Ser Leu Phe Ile Phe Arg Ser Gly
               200                 205                 210
Met Ala Pro Val Lys Tyr His Gly Asp Arg Ser Lys Glu Ser Leu
               215                 220                 225
Val Ser Phe Ala Met Gln His Val Arg Ser Thr Val Thr Glu Leu
               230                 235                 240
Trp Thr Gly Asn Phe Val Asn Ser Ile Gln Thr Ala Phe Ala Ala
               245                 250                 255
Gly Ile Gly Trp Leu Ile Thr Phe Cys Ser Lys Gly Gly Asp Cys
               260                 265                 270
Leu Thr Ser Gln Thr Arg Leu Arg Leu Ser Gly Met Leu Asp Gly
               275                 280                 285
```

```
Leu Val Asn Val Gly Trp Met Asp Cys Ala Thr Gln Asp Asn Leu
                290                 295                 300
Cys Lys Ser Leu Asp Ile Thr Thr Ser Thr Thr Ala Tyr Phe Pro
                305                 310                 315
Pro Gly Ala Thr Leu Asn Asn Lys Glu Lys Asn Ser Ile Leu Phe
                320                 325                 330
Leu Asn Ser Leu Asp Ala Lys Glu Ile Tyr Leu Glu Val Ile His
                335                 340                 345
Asn Leu Pro Asp Phe Glu Leu Leu Ser Ala Asn Thr Leu Glu Asp
                350                 355                 360
Arg Leu Ala His His Arg Trp Leu Leu Phe Phe His Phe Gly Lys
                365                 370                 375
Asn Glu Asn Ser Asn Asp Pro Glu Leu Lys Lys Leu Lys Thr Leu
                380                 385                 390
Leu Lys Asn Asp His Ile Gln Val Gly Arg Phe Asp Cys Ser Ser
                395                 400                 405
Ala Pro Asp Ile Cys Ser Asn Leu Tyr Val Phe Gln Pro Ser Leu
                410                 415                 420
Ala Val Phe Lys Gly Gln Gly Thr Lys Glu Tyr Glu Ile His His
                425                 430                 435
Gly Lys Lys Ile Leu Tyr Asp Ile Leu Ala Phe Ala Lys Glu Ser
                440                 445                 450
Val Asn Ser His Val Thr Thr Leu Gly Pro Gln Asn Phe Pro Ala
                455                 460                 465
Asn Asp Lys Glu Pro Trp Leu Val Asp Phe Phe Ala Pro Trp Cys
                470                 475                 480
Pro Pro Cys Arg Ala Leu Leu Pro Glu Leu Arg Arg Ala Ser Asn
                485                 490                 495
Leu Leu Tyr Gly Gln Leu Lys Phe Gly Thr Leu Asp Cys Thr Val
                500                 505                 510
His Glu Gly Leu Cys Asn Met Tyr Asn Ile Gln Ala Tyr Pro Thr
                515                 520                 525
Thr Val Val Phe Asn Gln Ser Asn Ile His Glu Tyr Glu Gly His
                530                 535                 540
His Ser Ala Glu Gln Ile Leu Glu Phe Ile Glu Asp Leu Met Asn
                545                 550                 555
Pro Ser Val Val Ser Leu Thr Pro Thr Thr Phe Asn Glu Leu Val
                560                 565                 570
Thr Gln Arg Lys His Asn Glu Val Trp Met Val Asp Phe Tyr Ser
                575                 580                 585
Pro Trp Cys His Pro Cys Gln Val Leu Met Pro Glu Trp Lys Arg
                590                 595                 600
Met Ala Arg Thr Leu Thr Gly Leu Ile Asn Val Gly Ser Ile Asp
                605                 610                 615
Cys Gln Gln Tyr His Ser Phe Cys Ala Gln Glu Asn Val Gln Arg
                620                 625                 630
Tyr Pro Glu Ile Arg Phe Phe Pro Pro Lys Ser Asn Lys Ala Tyr
                635                 640                 645
Gln Tyr His Ser Tyr Asn Gly Trp Asn Arg Asp Ala Tyr Ser Leu
                650                 655                 660
Arg Ile Trp Gly Leu Gly Phe Leu Pro Gln Val Ser Thr Asp Leu
                665                 670                 675
Thr Pro Gln Thr Phe Ser Glu Lys Val Leu Gln Gly Lys Asn His
                680                 685                 690
Trp Val Ile Asp Phe Tyr Ala Pro Trp Cys Gly Pro Cys Gln Asn
                695                 700                 705
Phe Ala Pro Glu Phe Glu Leu Leu Ala Arg Met Ile Lys Gly Lys
                710                 715                 720
Val Lys Ala Gly Lys Val Asp Cys Gln Ala Tyr Ala Gln Thr Cys
                725                 730                 735
Gln Lys Ala Gly Ile Arg Ala Tyr Pro Thr Val Lys Phe Tyr Phe
                740                 745                 750
Tyr Glu Arg Ala Lys Arg Asn Phe Gln Glu Glu Gln Ile Asn Thr
```

```
                755                 760                 765
Arg Asp Ala Lys Ala Ile Ala Ala Leu Ile Ser Glu Lys Leu Glu
                770                 775                 780
Thr Leu Arg Asn Gln Gly Lys Arg Asn Lys Asp Glu Leu
                785                 790
```

<210> 6
<211> 492
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3325534CD1

<400> 6

```
Met Ala Val Leu Leu Glu Thr Thr Leu Gly Asp Val Val Ile Asp
  1               5                  10                  15
Leu Tyr Thr Glu Glu Arg Pro Arg Ala Cys Leu Asn Phe Leu Lys
                20                  25                  30
Leu Cys Lys Ile Lys Tyr Tyr Asn Tyr Cys Leu Ile His Asn Val
                35                  40                  45
Gln Arg Asp Phe Ile Ile Gln Thr Gly Asp Pro Thr Gly Thr Gly
                50                  55                  60
Arg Gly Gly Glu Ser Ile Phe Gly Gln Leu Tyr Gly Asp Gln Ala
                65                  70                  75
Ser Phe Phe Glu Ala Glu Lys Val Pro Arg Ile Lys His Lys Lys
                80                  85                  90
Lys Gly Thr Val Ser Met Val Asn Asn Gly Ser Asp Gln His Gly
                95                 100                 105
Ser Gln Phe Leu Ile Thr Thr Gly Glu Asn Leu Asp Tyr Leu Asp
               110                 115                 120
Gly Val His Thr Val Phe Gly Glu Val Thr Glu Gly Met Asp Ile
               125                 130                 135
Ile Lys Lys Ile Asn Glu Thr Phe Val Asp Lys Asp Phe Val Pro
               140                 145                 150
Tyr Gln Asp Ile Arg Ile Asn His Thr Val Ile Leu Asp Asp Pro
               155                 160                 165
Phe Asp Asp Pro Pro Asp Leu Leu Ile Pro Asp Arg Ser Pro Glu
               170                 175                 180
Pro Thr Arg Glu Gln Leu Asp Ser Gly Arg Ile Gly Ala Asp Glu
               185                 190                 195
Glu Ile Asp Asp Phe Lys Gly Arg Ser Ala Glu Glu Val Glu Glu
               200                 205                 210
Ile Lys Ala Glu Lys Glu Ala Lys Thr Gln Ala Ile Leu Leu Glu
               215                 220                 225
Met Val Gly Asp Leu Pro Asp Ala Asp Ile Lys Pro Pro Glu Asn
               230                 235                 240
Val Leu Phe Val Cys Lys Leu Asn Pro Val Thr Thr Asp Glu Asp
               245                 250                 255
Leu Glu Ile Ile Phe Ser Arg Phe Gly Pro Ile Arg Ser Cys Glu
               260                 265                 270
Val Ile Arg Asp Trp Lys Thr Gly Glu Ser Leu Cys Tyr Ala Phe
               275                 280                 285
Ile Glu Phe Glu Lys Glu Glu Asp Cys Glu Lys Ala Phe Phe Lys
               290                 295                 300
Met Asp Asn Val Leu Ile Asp Asp Arg Arg Ile His Val Asp Phe
               305                 310                 315
Ser Gln Ser Val Ala Lys Val Lys Trp Lys Gly Lys Gly Gly Lys
               320                 325                 330
```

```
Tyr Thr Lys Ser Asp Phe Lys Glu Tyr Glu Lys Glu Gln Asp Lys
                335                 340                 345
Pro Pro Asn Leu Val Leu Lys Asp Lys Val Lys Pro Lys Gln Asp
                350                 355                 360
Thr Lys Tyr Asp Leu Ile Leu Asp Glu Gln Ala Glu Asp Ser Lys
                365                 370                 375
Ser Ser His Ser His Thr Ser Lys Lys His Lys Lys Lys Thr His
                380                 385                 390
His Cys Ser Glu Glu Lys Glu Asp Glu Asp Tyr Met Pro Ile Lys
                395                 400                 405
Asn Thr Asn Gln Asp Ile Tyr Arg Glu Met Gly Phe Gly His Tyr
                410                 415                 420
Glu Glu Glu Glu Ser Cys Trp Glu Lys Gln Lys Ser Glu Lys Arg
                425                 430                 435
Asp Arg Thr Gln Asn Arg Ser Arg Ser Arg Ser Arg Glu Arg Asp
                440                 445                 450
Gly His Tyr Ser Asn Ser His Lys Ser Lys Tyr Gln Thr Asp Leu
                455                 460                 465
Tyr Glu Arg Glu Arg Ser Lys Lys Arg Asp Arg Ser Arg Ser Pro
                470                 475                 480
Lys Lys Ser Lys Asp Lys Glu Lys Ser Lys Tyr Arg
                485                 490
```

<210> 7
<211> 160
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3817050CD1

<400> 7

```
Met Val Ile Pro Thr Val Pro Phe Asn Ile Thr Ile Asn Ser Lys
  1                 5                  10                  15
Pro Leu Gly His Ile Ser Phe Gln Leu Phe Ala Asp Lys Phe Pro
                 20                  25                  30
Lys Thr Gly Glu Asn Phe His Thr Leu Asn Asn Lys Asp Lys Gly
                 35                  40                  45
Phe Gly Ser Cys Phe His Arg Ile Ile Pro Glu Phe Ile Cys Gln
                 50                  55                  60
Gly Asp Asp Phe Thr Pro His Asn Gly Ile Gly Gly Lys Ser Ile
                 65                  70                  75
Tyr Gly Asp Lys Phe Asp Asp Lys Asn Phe Ile Val Lys His Thr
                 80                  85                  90
Gly Leu Gly Ile Leu Ser Met Ala Asn Ala Ala Pro Lys Thr Asn
                 95                 100                 105
Glu Ser Gln Phe Phe Ile Cys Thr Ala Met Ala Lys Trp Trp Asp
                110                 115                 120
Gly Lys His Val Ile Phe Gly Arg Val Lys Glu Gly Met Asn Ile
                125                 130                 135
Val Glu Ala Met Glu Cys Phe Gly Ser Arg Asn Gly Lys Thr Ser
                140                 145                 150
Lys Ile Ala Ile Ala Asn Cys Arg Gln Leu
                155                 160
```

<210> 8

<211> 744
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5324378CD1

<400> 8

```
Met Gln Lys Thr Glu Thr Leu Leu Leu Phe Ser Cys Asn Ile Ser
  1               5                  10                  15
Val Ser Ser Glu Pro Gly Val Leu Gly Tyr Phe Glu Phe Ser Gly
             20                  25                  30
Ser Pro Gln Pro Pro Gly Tyr Leu Thr Phe Phe Thr Ser Ala Leu
             35                  40                  45
His Ser Leu Lys Lys Asp Tyr Leu Gly Thr Val Arg Phe Gly Val
             50                  55                  60
Ile Thr Asn Lys His Leu Ala Lys Leu Val Ser Leu Val His Ser
             65                  70                  75
Gly Ser Val Tyr Leu His Arg His Phe Asn Thr Ser Leu Val Phe
             80                  85                  90
Pro Arg Glu Val Leu Asn Tyr Thr Ala Glu Asn Ile Cys Lys Trp
             95                 100                 105
Ala Leu Glu Asn Gln Glu Thr Leu Phe Arg Trp Leu Arg Pro His
            110                 115                 120
Gly Gly Lys Ser Leu Leu Leu Asn Asn Glu Leu Lys Lys Gly Pro
            125                 130                 135
Ala Leu Phe Leu Phe Ile Pro Phe Asn Pro Leu Ala Glu Ser His
            140                 145                 150
Pro Leu Ile Asp Glu Ile Thr Glu Val Ala Leu Glu Tyr Asn Asn
            155                 160                 165
Cys His Gly Asp Gln Val Val Glu Arg Leu Leu Gln His Leu Arg
            170                 175                 180
Arg Val Asp Ala Pro Val Leu Glu Ser Leu Ala Leu Glu Val Pro
            185                 190                 195
Ala Gln Leu Pro Asp Pro Pro Thr Ile Thr Ala Ser Pro Cys Cys
            200                 205                 210
Asn Thr Val Val Leu Pro Gln Trp His Ser Phe Ser Arg Thr His
            215                 220                 225
Asn Val Cys Glu Leu Cys Val Asn Gln Thr Ser Gly Gly Met Lys
            230                 235                 240
Pro Ser Ser Val Ser Val Pro Gln Cys Ser Phe Phe Glu Met Ala
            245                 250                 255
Ala Ala Leu Asp Ser Phe Tyr Leu Lys Glu Gln Thr Phe Tyr His
            260                 265                 270
Val Ala Ser Asp Ser Ile Glu Cys Ser Asn Phe Leu Thr Ser Tyr
            275                 280                 285
Ser Pro Phe Ser Tyr Tyr Thr Ala Cys Cys Arg Thr Ile Ser Arg
            290                 295                 300
Gly Val Ser Gly Phe Ile Asp Ser Glu Gln Gly Val Phe Glu Ala
            305                 310                 315
Pro Thr Val Ala Phe Ser Ser Leu Glu Lys Lys Cys Glu Val Asp
            320                 325                 330
Ala Pro Ser Ser Val Pro His Ile Glu Glu Asn Arg Tyr Leu Phe
            335                 340                 345
Pro Glu Val Asp Met Thr Ser Thr Asn Phe Thr Gly Leu Ser Cys
            350                 355                 360
Arg Thr Asn Lys Thr Leu Asn Ile Tyr Leu Leu Asp Ser Asn Leu
            365                 370                 375
Phe Trp Leu Tyr Ala Glu Arg Leu Gly Ala Pro Ser Ser Thr Gln
            380                 385                 390
Val Lys Glu Phe Ala Ala Ile Val Asp Val Lys Glu Glu Ser His
            395                 400                 405
```

```
Tyr Ile Leu Asp Pro Lys Gln Ala Leu Met Lys Leu Thr Leu Glu
                410                 415                 420
Ser Phe Ile Gln Asn Phe Ser Val Leu Tyr Ser Pro Leu Lys Arg
                425                 430                 435
His Leu Ile Gly Ser Gly Ser Ala Gln Phe Pro Ser Gln His Leu
                440                 445                 450
Ile Thr Glu Val Thr Thr Asp Thr Phe Trp Glu Val Val Leu Gln
                455                 460                 465
Lys Gln Asp Val Leu Leu Leu Tyr Tyr Ala Pro Trp Cys Gly Phe
                470                 475                 480
Cys Pro Ser Leu Asn His Ile Phe Ile Gln Leu Ala Arg Asn Leu
                485                 490                 495
Pro Met Asp Thr Phe Thr Val Ala Arg Ile Asp Val Ser Gln Asn
                500                 505                 510
Asp Leu Pro Trp Glu Phe Met Val Asp Arg Leu Pro Thr Val Leu
                515                 520                 525
Phe Phe Pro Cys Asn Arg Lys Asp Leu Ser Val Lys Tyr Pro Glu
                530                 535                 540
Asp Val Pro Ile Thr Leu Pro Asn Leu Leu Arg Phe Ile Leu His
                545                 550                 555
His Ser Asp Pro Ala Ser Ser Pro Gln Asn Val Ala Asn Ser Pro
                560                 565                 570
Thr Lys Glu Cys Leu Gln Ser Glu Ala Val Leu Gln Arg Gly His
                575                 580                 585
Ile Ser His Leu Glu Arg Glu Ile Gln Lys Leu Arg Ala Glu Ile
                590                 595                 600
Ser Ser Leu Gln Arg Ala Gln Val Gln Val Glu Ser Gln Leu Ser
                605                 610                 615
Ser Ala Arg Arg Asp Glu His Arg Leu Arg Gln Gln Gln Arg Ala
                620                 625                 630
Leu Glu Glu Gln His Ser Leu Leu His Ala His Ser Glu Gln Leu
                635                 640                 645
Gln Ala Leu Tyr Glu Gln Lys Thr Arg Glu Leu Gln Glu Leu Ala
                650                 655                 660
Arg Lys Leu Gln Glu Leu Ala Asp Ala Ser Glu Asn Leu Leu Thr
                665                 670                 675
Glu Asn Thr Trp Leu Lys Ile Leu Val Ala Thr Met Glu Arg Lys
                680                 685                 690
Leu Glu Gly Arg Asp Gly Ala Glu Ser Leu Ala Ala Gln Arg Glu
                695                 700                 705
Val His Pro Lys Gln Pro Glu Pro Ser Ala Thr Pro Gln Leu Pro
                710                 715                 720
Gly Ser Ser Pro Pro Pro Ala Asn Val Ser Ala Thr Leu Val Ser
                725                 730                 735
Glu Arg Asn Lys Glu Asn Arg Thr Asp
                740
```

<210> 9
<211> 2015
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 011886CB1

<400> 9
agacgttgtt gcttgggcgc ttctccgctg cgtgtaggtg aagggggctt cctgaccgag 60
acatggattt aggtgctatt acaaaatact cagcattaca cgccaagccc aatggactga 120
tccttcaata cgggactgct ggatttcgaa cgaaggcaga acatcttgat catgtcatgt 180

ttcgcatggg attattagct gtcctgaggt caaaacagac aaaatccact ataggagtca 240
tggtaacagc gtcccacaat cctgaggaag acaatggtgt aaaattggtt gatcctttgg 300
gtgaaatgtt ggcaccatcc tgggaggaac atgccacctg tttagcaaat gctgaggaac 360
aagatatgca gagagtgctt attgacatca gcgagaaaga agctgtgaat ctgcaacaag 420
atgcctttgt agttattggt agagatacca ggcccagcag tgagaaactt tcacaatctg 480
taatagatgg tgtgactgtt ctaggaggtc aattccatga ttatggcttg ttaacaacac 540
cccagctgca ctacatggtg tattgtcgaa acacgggtgg ccgatatgga aaggcaacta 600
tagaaggtta ctaccagaaa ctctctaagg cttttgtgga actcaccaaa caggcttctt 660
gcagtggaga tgaatacaga tcacttaagg ttgactgtgc aaatggcata ggggccctga 720
agctaaggga aatggaacac tacttctcac agggcctgtc agttcagctg tttaatgatg 780
ggtccaaggg caaactcaat catttatgtg gagctgactt tgtgaaaagt catcagaaac 840
ctccacaggg aatggaaatt aagtccaatg aaagatgctg ttcttttgat ggagatgcag 900
acagaattgt ttattactac catgatgcag atggccactt tcatctcata gatggagaca 960
agatagcaac gttaattagc agtttcctta aagagctcct ggtggagatt ggagaaagtt 1020
tgaatattgg tgttgtacaa actgcatatg caaatggaag ttcaacacgg tatcttgaag 1080
aagttatgaa ggtacctgtc tattgcacta agactggtgt aaaacatttg caccacaagg 1140
ctcaagagtt tgacattgga gtttattttg aagcaaatgg gcatggcact gcactgttta 1200
gtacagctgt tgaaatgaag ataaaacaat cagcagaaca actggaagat aagaaaagaa 1260
aagctgctaa gatgcttgaa aacattattg acttgtttaa ccaggcagct ggtgatgcta 1320
tttctgacat gctggtgatt gaagcaatct tggctctgaa gggcttgact gtacaacagt 1380
gggatgctct ctatacagat cttccaaaca gacaacttaa agttcaggtt gcagacagga 1440
gagttattag cactaccgat gctgaaagac aagcagttac acccccagga ttacaggagg 1500
caatcaatga cctggtgaag aagtacaagc tttctcgagc ttttgtccgg ccctctggta 1560
cagaagatgt cgtccgagta tatgcagaag cagactcaca agaaagtgca gatcaccttg 1620
cacatgaagt gagcttggca gtatttcagc tggctggagg aattggagaa aggccccaac 1680
caggtttctg aagataattt tcatattcct gagaaactgg acttttaca agtctttaca 1740
aaactgtcaa taataatggc agtactaaga gatttataat cataatgttt acaatgcagc 1800
ctactggatt gtctctagat ctgtttttct taaacactaa cagaataatt ctttataaat 1860
aggtaagcct tacacttgtt aaagaaattt acctctaatt tcagtctcac taatgtaaaa 1920
tactgggact taagtataca attcagtcac taactgtaca gttttatgtg gggaacaatt 1980
catgcaggct actggaaaat taaatcttat tacca 2015

<210> 10
<211> 1823
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1863189CB1

<400> 10
cggctcgagc cggagcgagg cgtcgggatg cagcgccccg ggcccttcag caccctctac 60
gggcgggtct tggccccgct gcccgggagg gccgggggcg cggcctccgg cggaggaggg 120
aacagctggg acctcccggg ttcccacgtg cggctgccgg ggcgtgcaca gtctgggacc 180
cgtggcggtg ctggcaacac aagcaccagc tgcggggact ccaacagcat ctgcccggcc 240
ccctccacga tgtccaaggc cgaggaggcc aagaagctgg cgggccgcgc ggctgtggag 300
aaccacgtga ggaataacca agtgctggga attggaagtg gttctacaat tgtccatgct 360
gtgcagcgaa tagctgaaag ggtgaagcaa gagaatctga acctcgtctg tattcccact 420
tccttccagg cccgccagct catcctgcag tatggcttga ccctcagtga tctggatcga 480
cacccagaga tcgaccttgc catcgatggt gctgatgaag tagatgctga tctcaatctc 540
atcaagggtg gcggaggctg cctgacccag gagaagattg tggctggcta tgctagtcgc 600
ttcatcgtga tcgctgattt caggaaagat tcgaagaatc tcggggatca gtggcacaag 660
ggaatcccca tcgaggtcat cccaatggcc tatgtcccag tgagccgagc tgtgagccag 720
aagtttgggg gcgtggttga acttcgaatg gctgtcaaca aggctggtcc tgtggtgaca 780
gataatggga attttatctt ggactggaag tttgaccggg tacacaaatg gagtgaagtg 840
aatacagcta tcaaaatgat cccaggtgtg gtggacacag gcctattcat caacatggct 900
gagagagtct actttgggat gcaggatggc tcagtgaaca tgagggagaa gcctttctgt 960
tgaccctgca aggagcagag tgtgttcacc ttgagtctcc agcccacagc caaggtggac 1020
gtacctctcc aggagccttt gccttaatgt atctctgcct ggacaacttg tggtgggggg 1080
tgggggaag agtgggaggg ggagttaaat ccagtcttat gaagtattgt tattaaatgt 1140
ctttttaaaa agagaaatat aaacatatat ttttactatt aaaatattca gtttttaaa 1200

tgaagtagaa cttgagttca tgttttatat gaaatattta ccaaaaaaaa aaaatgaggt 1260
aaactgtatt taaaaccttt gacttgagtc tgctggtaaa gcttctgaat attgagtttg 1320
ctgagaaata aaaatcaaaa cttctttaag ctggtaaagt gaggggccca ccagcagtga 1380
tctcctgatg ccttactgga aactttgttt acttgtctgc taccctctga tttgttttta 1440
gttagttttt attgtgagca cacatagtac ctagttacat cttaagatca ggtttataaa 1500
actgtggagt ggagcggtat ggtatggaat gacttggaat gtaagctgtc agggagaaaa 1560
tgttgttaca cttttgctaa gatctggggg tttcttcata ttcctgctgt tggaagcagt 1620
tgaccagaaa tgcttgccag tactgccaaa gcactgctgt gaaatgtgaa gtactttgtt 1680
tttttatttt taatgatttt ctttttgtta ttaatatttt tctctgttcc tttgttatta 1740
cttgcatggt ttggcgtcag aagtccttac ctctttatat tgtttgcagg tttaaataaa 1800
acagtgtggt gccaaaaaaa aaa 1823

<210> 11
<211> 1526
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2088868CB1

<400> 11
gcaggagcag gagagggaca atggaagctg ccccgtccag gttcatgttc ctcttatttc 60
tcctcacgtg tgagctggct gcagaagttg ctgcagaagt tgagaaatcc tcagatggtc 120
ctggtgctgc ccaggaaccc acgtggctca cagatgtccc agctgccatg gaattcattg 180
ctgccactga ggtggctgtc ataggcttct tccaggattt agaaatacca gcagtgccca 240
tactccatag catggtgcaa aaattcccag gcgtgtcatt tgggatcagc actgattctg 300
aggttctgac acactacaac atcactggga acaccatctg cctctttcgc ctggtagaca 360
atgaacaact gaatttagag gacgaagaca ttgaaagcat tgatgccacc aaattgagcc 420
gtttcattga gatcaacagc ctccacatgg tgacagagta caaccctgtg actgtgattg 480
ggttattcaa cagcgtaatt cagattcatc tcctcctgat aatgaacaag gcctccccag 540
agtatgaaga gaacatgcac agataccaga aggcagccaa gctcttccag gggaagattc 600
tctttattct ggtggacagt ggtatgaaag aaaatgggaa ggtgatatca tttttcaaac 660
taaaggagtc tcaactgcca gctttggcaa tttaccagac tctagatgac gagtgggata 720
cactgcccac agcagaagtt tccgtagagc atgtgcaaaa cttttgtgat ggattcctaa 780
gtggaaaatt gttgaaagaa aatcgtgaat cagaaggaaa gactccaaag gtggaactct 840
gacttctcct tggaactaca tatggccaag tatctacttt atgcaaagta aaaaggcaca 900
actcaaatct cagagacact aaacaacagg atcactaggc ctgccaacca cacacacacg 960
cacgtgcaca cacgcacgca cgcgtgcaca cacacacgcg cacacacaca cacacacaca 1020
cagagcttca tttcctgtct taaaatctcg ttttctcttc ttccttcttt taaatttcat 1080
atcctcactc cctatccaat ttccttctta tcgtgcattc atactctgta agcccatctg 1140
taacacacct agatcaaggc tttaagagac tcactgtgat gcctctatga aagagaggca 1200
ttcctagaga aagattgttc caatttgtca tttaatatca agtttgtata ctgcacatga 1260
cttacacaca acatagttcc tgctctttta aggttaccta agggttgaaa ctctaccttc 1320
tttcataagc acatgtccgt ctctgactca ggatcaaaaa ccaaaggatg gttttaaaca 1380
cctttgtgaa attgtctttt tgccagaagt taaaggctgt ctccaagtcc ctgaactcag 1440
cagaaataga ccatgtgaaa actccatgct tggttagcat ctccaactcc ctatgtaaat 1500
caacaacctg cataataaat aacaga 1526

<210> 12
<211> 1205
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2481256CB1

<400> 12
gcggtatggc gtcgggctgc aagattggcc cgtccatcct caacagcgac ctggccaatt 60
taggggccga gtgcctccgg atgctagact ctggggccga ttatctgcac ctggacgtaa 120

tggacgggca ttttgttccc aacatcacct ttggtcaccc tgtggtagaa agccttcgaa 180
agcagctagg ccaggaccct ttctttgaca tgcacatgat ggtgtccaag ccagaacagt 240
gggtaaagcc aatggctgta gcaggagcca atcagtacac ctttcatctc gaggctactg 300
agaacccagg ggctttgatt aaagacattc gggagaatgg gatgaaggtt ggccttgcca 360
tcaaaccagg aacctcagtt gagtatttgg caccatgggc taatcagata gatatggcct 420
tggttatgac agtggaaccg gggtttggag ggcagaaatt catggaagat atgatgccaa 480
aggttcactg gttgaggacc cagttcccat ctttggatat agaggtcgat ggtggagtag 540
gtcctgacac tgtccataaa tgtgcagagg caggagctaa catgattgtg tctggcagtg 600
ctattatgag gagtgaagac cccagatctg tgatcaatct attaagaaat gtttgctcag 660
aagctgctca gaaacgttct cttgatcggt gaaaccataa ggagcccagt gttcctgttc 720
atgaaatctc ccttttactg gaaaacagga atattgacta ccaaatcaca atgcaattga 780
agccgtactg ctttttgag cagttattca ttccagtgat taaaactgat tgtgcagaat 840
attctaagag gtcagaaatt ggtgtgtata actacatttt tagtgatgca atttattgat 900
tagtgagtaa gatactgttt ttattgagag atttgatttt tataaagtaa aaatacggct 960
gcattagggt tacaaacaga aaagtgtctt aatgtctaag gagggcatat tagctacact 1020
acaaaaacaa attttgtctg tacttctgaa aagaattttg ttgtttctca gctgttttcc 1080
aaaagcaaag gaagtcttta tggttttttt ctatttcatg ttattgtgat ttgtttataa 1140
gtttgggtgg ggtgcatacc atattcttgg ttcttaaaat ctatcacttt tcaccttata 1200
cttga 1205

<210> 13
<211> 4796
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2505257CB1

<400> 13
gccccggctc gccgtggaga ccggcgcgtg aggcacctac cggtaccggc cgcgcgctgg 60
tagtcgccgg tgtggctgca cctcaccaat cccgtgcgcc gcggctgggc cgtcggagag 120
tgcgtgtgct tctctcctgc acgcggtgct tgggctcggc caggcggggt ccgccgccag 180
ggtttgagga tgggggagta gctacaggaa gcgaccccgc gatggcaagg tatatttttg 240
tggaatgaaa aggaagtatt agaaatgagc tgaagaccat tcacagatta atatttttgg 300
ggacagattt gtgatgcttg attcaccctt gaagtaatgt agacagaagt tctcaaattt 360
gcatattaca tcaactggaa ccagcagtga atcttaatgt tcacttaaat cagaacttgc 420
ataagagaga gaatgggagt ctggttaaat aaagatgacg atatcagaga cttgaaaagg 480
atcattctct gttttctgat agtgtatatg gccattttag tgggcacaga tcaggatttt 540
tacagtttac ttggagtgtc caaaactgca agcagtagag aaataagaca agctttcaag 600
aaattggcat tgaagttaca tcctgataaa aacccgaata acccaaatgc acatggcaat 660
tttttaaaaa taaatagagc atatgaagta ctcaaagatg aagatctacg gaaaaagtat 720
gacaaatatg gagaaaaggg acttgaggat aatcaaggtg gccagtatga aagctggaac 780
tattatcgtt atgattttgg tatttatgat gatgatcctg aaatcataac attggaaaga 840
agagaatttg atgctgctgt taattctgga gaactgtggt ttgtaaattt ttactcccca 900
ggctgttcac actgccatga tttagctccc acatggagag actttgctaa agaagtggat 960
gggttacttc gaattggagc tgttaactgt ggtgatgata gaatgctttg ccgaatgaaa 1020
ggagtcaaca gctatcccag tctcttcatt tttcggtctg gaatggcccc agtgaaatat 1080
catggagaca gatcaaagga gagtttagtg agttttgcaa tgcagcatgt tagaagtaca 1140
gtgacagaac tttggacagg aaattttgtc aactccatac aaactgcttt tgctgctggt 1200
attggctggc tgatcacttt ttgttcaaaa ggaggagatt gtttgacttc acagacacga 1260
ctcaggctta gtggcatgtt ggatggtctt gttaatgtag gatggatgga ctgtgccacc 1320
caggataacc tttgtaaaag cttagatatt acaacaagta ctactgctta ttttcctcct 1380
ggagccactt taaataacaa agagaaaaac agtattttgt ttctcaactc attggatgct 1440
aaagaaatat atttggaagt aatacataat cttccagatt ttgaactact ttcggcaaac 1500
acactagagg atcgtttggc tcatcatcgg tggctgttat tttttcattt tggaaaaaat 1560
gaaaattcaa atgatcctga gctgaaaaaa ctaaaaactc tacttaaaaa tgatcatatt 1620
caagttggca ggtttgactg ttcctctgca ccagacatct gtagtaatct gtatgttttt 1680
cagccgtctc tagcagtatt taaaggacaa ggaaccaaag aatatgaaat tcatcatgga 1740
aagaagattc tatatgatat acttgccttt gccaaagaaa gtgtgaattc tcatgttacc 1800
acgcttggac ctcaaaattt tcctgccaat gacaaagaac catggcttgt tgatttcttt 1860
gccccctggt gtccaccatg tcgagcttta ctaccagagt tacgaagagc atcaaatctt 1920

ctttatggtc agcttaagtt tggtacacta gattgtacag ttcatgaggg actctgtaac 1980
atgtataaca ttcaggctta tccaacaaca gtggtattca accagtccaa cattcatgag 2040
tatgaaggac atcactctgc tgaacaaatc ttggagttca tagaggatct tatgaatcct 2100
tcagtggtct cccttacacc caccaccttc aacgaactag ttacacaaag aaaacacaac 2160
gaagtctgga tggttgattt ctattctccg tggtgtcatc cttgccaagt cttaatgcca 2220
gaatggaaaa gaatggcccg gacattaact ggactgatca acgtgggcag tatagattgc 2280
caacagtatc attctttttg tgcccaggaa aacgttcaaa gataccctga gataagattt 2340
tttcccccaa aatcaaataa agcttatcag tatcacagtt acaatggttg gaatagggat 2400
gcttattccc tgagaatctg gggtctagga tttttacctc aagtatccac agatctaaca 2460
cctcagactt tcagtgaaaa agttctacaa gggaaaaatc attgggtgat tgatttctat 2520
gctccttggt gtggaccttg ccagaatttt gctccagaat ttgagctctt ggctaggatg 2580
attaaaggaa aagtgaaagc tggaaaagta gactgtcagg cttatgctca gacatgccag 2640
aaagctggga tcagggccta tccaactgtt aagttttatt tctacgaaag agcaaagaga 2700
aattttcaag aagagcagat aaataccaga gatgcaaaag caatcgctgc cttaataagt 2760
gaaaaattgg aaactctccg aaatcaaggc aagaggaata aggatgaact ttgataatgt 2820
tgaagatgaa gaaaaagttt aaaagaaatt ctgacagatg acatcagaag acacctattt 2880
agaatgttac atttatgatg ggaatgaatg aacattatct tagacttgca gttgtactgc 2940
cagaattatc tacagcactg gtgtaaaaga agggtctgca aacttttct gtaaagggcc 3000
ggtttataaa tattttagac tttgcaggct ataatatatg gttcacacat gagaacaaga 3060
atagagtcat catgtattct ttgttatttg cttttaacaa cctttaaaaa atattaaaac 3120
gattcttagc tcagagccat acaaaagtag gctggattca gtccatggac catagattgc 3180
tgtccccctc gacggactta taatgtttca ggtggctggc ttgaacatga gtctgctgtg 3240
ctatctacat aaatgtctaa gttgtataaa gtccactttc ccttcacgtt ttttggctga 3300
cctgaaaaga ggtaacttag tttttggtca cttgttctcc taaaaatgct atccctaacc 3360
atatatttat atttcgtttt aaaaacaccc atgatgtggc acagtaaaca aaccctgtta 3420
tgctgtatta ttatgaggag attcttcatt gttttctttc cttctcaaag gttgaaaaaa 3480
tgcttttaat ttttcacagc cgagaaacag tgcagcagta tatgtgcaca cagtaagtac 3540
acaaatttga gcaacagtaa gtgcacaaat tctgtagttt gctgtatcat ccaggaaaac 3600
ctgagggaaa aaaattatag caattaactg ggcattgtag agtatcctaa atatgttatc 3660
aagtatttag agttctatat tttaaagata tatgtgttca tgtattttct gaaattgctt 3720
tcatagaaat tttcccactg atagttgatt tttgaggcat ctaatattta catatttgcc 3780
ttctgaactt tgttttgacc tgtatccttt atttacattg ggtttttctt tcgtagtttt 3840
ggtttttcac tcctgtccag tctatttatt attcaaatag gaaaaattac tttacaggtt 3900
gttttactgt agcttataat gatactgtag ttattccagt tactagttta ctgtcagagg 3960
gctgcctttt tcagataaat attgacataa taactgaagt tatttttata agaaaatcaa 4020
gtatataaat ctaggaaagg gatcttctag tttctgtgtt gtttagactc aaagaatcac 4080
aaatttgtca gtaacatgta gttgtttagt tataattcag agtgtacaga atggtaaaaa 4140
ttccaatcag tcaaaagagg tcaatgaatt aaaaggcttg caacttttct caaaaacctg 4200
ttagaatatg ctttattgtg ttttgaggag ttttcctttt tttcttttca atatcacttt 4260
atcctccagt atttcctcat aagggttatt atagccataa ttaatgttaa aatagacttt 4320
gttcttcata ttctcccatc tttttcgcta ctatatactc tgtctggatt ctgctgtatg 4380
cctgttggca tatatggaac agtcaccact tgtcacactt aacaccagct ttttgaatta 4440
tgatcagtaa tggcaagagc ctttcattct cgaatgttta aagcctagga gttctacaaa 4500
attggcttct ttctacaaga atcccaaaat ggaatgccta aagaagtctt acttgggtaa 4560
atacttacta aaatatactg gttatgtgca tatcaccaca ctggacactg aggagtgttc 4620
aaaaggaatc taagacatgg tccccatctt ccaactgtct gtaattcact gttttgtcat 4680
tgagctcata aggtacttac attactacct ataaatgttt cctgtacttg ttagttgttg 4740
agaaacattt taggcagtaa ataaaatagt aaatattatg tgtcctataa tttgac 4796

<210> 14
<211> 1680
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3325534CB1

<400> 14
acacccccct cctcccgggg tttgtagcgg aggaggagcg ggcgccatgg cggttctact 60
ggagaccact ttaggcgacg tcgtcatcga cttgtacacc gaagaacggc cgcgtgcctg 120
cttgaatttc ttgaaactgt gcaaaataaa atattacaat tattgcctta ttcacaatgt 180

acagagggat tttatcatac aaactggcga tcctacaggg actggccgtg gaggagagtc 240
tatctttggc caactgtatg gtgatcaagc aagctttttt gaggcagaaa aagtcccaag 300
aattaagcac aagaagaaag gcacagtgtc catggtgaat aatggcagtg atcaacatgg 360
atctcagttt cttatcacca caggagaaaa tctagattat cttgatggtg tccatacggt 420
gtttggtgag gtgacagaag gcatggacat aattaagaaa attaatgaga cctttgttga 480
caaggacttt gtaccatatc aggatatcag gataaatcat acggtgattt tagatgatcc 540
atttgatgac cctcctgatt tattaatccc tgatcgatca ccagaaccta caagggaaca 600
attagatagt ggtcgaatag gagcagatga agaaattgat gatttcaaag gaagatcagc 660
tgaggaagta gaagaaataa aggcagaaaa agaggctaaa actcaggcta tacttttgga 720
gatggtggga gacctacctg atgcagatat taaacctcca gaaaatgtac tgtttgtgtg 780
taaattgaac ccagtgacca cagatgagga tctggaaata atattctcta gatttgggcc 840
aataagaagt tgtgaagtta tccgagactg gaagacagga gagtccctct gttacgcttt 900
tattgaattt gaaaaggaag aagattgtga gaaagcattc ttcaaaatgg acaatgtgct 960
tatagatgac agaagaatac atgtggattt tagccagtcg gttgcaaagg ttaaatggaa 1020
aggaaaaggt gggaaataca ccaagagtga tttcaaggag tatgaaaaag aacaggataa 1080
accacctaat ttggttctga aagataaagt aaagcccaaa caggatacaa aatacgatct 1140
tatattagat gagcaggccg aagactcaaa atcaagtcac tcacacacaa gtaaaaaaca 1200
caagaagaaa acccatcact gttctgaaga gaaagaagat gaggactaca tgccaatcaa 1260
aaatactaat caggatatct atagagaaat ggggtttggt cactatgaag aagaagaaag 1320
ctgttgggag aaacaaaaga gtgaaaagag agaccgaact cagaaccgaa gtcgtagccg 1380
atctcgagag agggatggcc attatagtaa tagtcataaa tcaaaatacc aaacagatct 1440
ttatgaaaga gaaaggagta aaaagagaga ccgaagcaga agtccaaaga agtccaaaga 1500
taaagaaaaa tctaagtata gatgaaagat gaagaggcag aattgagagg ctaacatatt 1560
tactcttgtc taacttaaga gtgccaggaa agcagatgct tagattttgt gtcaaagctt 1620
gttatttttt tcatactagg attatggtct ttagattaat actgattata tagagcacgc 1680

<210> 15
<211> 1403
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3817050CB1

<400> 15
cctgtgcact gttggtggga atataaaatg atgcagctgg ctttgcagac actgctgtcc 60
cccaacaccc cctgtcacta ggccatggtc atcccgactg tgcccttcaa catcaccatc 120
aacagcaagc ccttaggaca catctccttt cagctatttg cagacaaatt tccaaagaca 180
ggagaaaact ttcacactct gaacaataaa gacaaaggat ttggttcctg ctttcacaga 240
attattccgg agtttatatg ccagggtgat gacttcacac cccataatgg cattggtggc 300
aagtccatct acggggataa atttgatgat aagaacttta ttgtgaagca tacaggtctt 360
ggcatcttgt ccatggcaaa tgctgcaccc aaaacaaatg agtcccagtt tttcatctgc 420
actgccatgg ccaaatggtg ggatggcaag catgtgatct ttggcagggt gaaagagggc 480
atgaatattg tggaagccat ggaatgcttt gggtccagga atggcaagac aagcaagatc 540
gccattgcca actgcagaca actctgataa atttgacttg tgttttatct taaccaccag 600
acctttcctt ttgtagctca ggagagcacc gttccacccc attcgctcac aatatcctat 660
aatatttgtg ctctcactgc agttctttga gttctatatt ttcattatgt ccctccacgt 720
atagctggat tgcagagtta agtttatgat tatgaaataa aaactaacaa aaaaaaatga 780
tgcagccact atggaaaaca gtatcacagt ttctcaaata attaaacatt gaattactat 840
atgattcagc agttccactc ctggatatat atccaaaaga attgaaagca gaattccaaa 900
gaaatatttg cacatccatg ttcatagcta taccattcac agtagccaag aggtggaagc 960
catctgtgtg cccatccaca gatgaatgga taaacaaaat atgggatata cacactatga 1020
atacagcctt aaaaaggaag gaaattccaa cacatgctac aacatggagg aatcttgagg 1080
aattaacggt aagtgaaata agccagtcac aaaaaggcca atactgaatg attccactta 1140
tgtgaggtat ctagagtagt catattcata gagacagaaa atagaatgat tgttgccagc 1200
aactgggagg aaggggtgtg gaaaagttgt ttaatggata ttgagtttgt tttcccagac 1260
gaagaagttc tgaaggttgg ttacatgatg tgaatatact aaacactact gaactgtgta 1320
cttagaatgg ttaagataaa ttttatgcgt tttcactaca ataacaagta gaacagtaga 1380
acagatgatt agtcacagca gaa 1403

<210> 16
<211> 2665
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5324378CB1

<400> 16
ccgtggcgct cggctgcgcg ctgctcctcg ccctcaagtt cacctgcagt cgagcaaaag 60
atgtgataat accagcaaag ccacctgtca gctttttctc cttgaggtct ccagtccttg 120
acctcttcca ggggcagctg gattatgcag agtacgttcg acgggattca gaggtggtac 180
tgctcttctt ctatgccccc tggtgtggac agtccatcgc tgccagggca gaaattgagc 240
aagcagcaag tcggctttca gatcaggtgt tgtttgtggc aattaactgt tggtggaacc 300
aggggaaatg cagaaaacag aaacacttct tttattttcc tgtaatatat ctgtatcatc 360
ggagcctgga gtactcgggt actttgagtt cagtggctca ccccagcctc ctggttattt 420
gaccttcttc acctcagcat tacattcatt aaagaaagat tacctaggaa cagtacgatt 480
tggggttatc acaaataaac atcttgcgaa actggtatcc ttagtacact ctggaagtgt 540
gtatttacat agacatttca acacatcact tgtcttcccc agggaggtcc tgaactacac 600
agctgagaac atctgtaagt gggccttaga aaaccaggag acgctctttc ggtggctgcg 660
gccacacgga ggcaagagtc tcctgctgaa taacgagctg aagaaaggac cagcgctgtt 720
tctgttcata ccttttaatc ccctggccga aagtcatcct ttaatagacg agatcaccga 780
agtggccttg gagtacaaca actgtcatgg ggaccaggtg gtggagcgtc tccttcagca 840
cctgcggcgg gtggatgctc cagtgctgga gtccctggcc ctggaagtgc cggcacagct 900
gccagacccg ccaacgatca cagcgtcccc ctgctgcaac actgtggtgc tgccccagtg 960
gcactccttc tccaggaccc acaacgtctg tgaactctgt gtcaaccaga cctccggggg 1020
catgaagccg agctcggtca gcgtgccaca gtgcagcttt tttgaaatgg cagcagctct 1080
ggattctttc tacctcaagg agcagacctt ttatcatgtg gcatcagaca gcatagaatg 1140
cagcaatttt ttaacttcct atagcccctt cagctactac actgcatgtt gcaggaccat 1200
aagcaggggt gtgtcaggct tcatcgactc tgaacaaggt gtctttgaag cccctactgt 1260
tgcattttct tcccttgaga agaaatgtga ggttgatgcc ccaagctccg ttcctcacat 1320
tgaggagaac aggtatctct ttccagaagt ggacatgact agcacaaact tcacaggcct 1380
gagctgcaga accaacaaga ctctcaacat ctacctttg gattcaaatt tgttttggtt 1440
atatgcagag agactgggtg ctccgagctc cactcaggtg aaagaatttg cggcaattgt 1500
tgacgtgaaa gaagaatctc attacatctt ggatccaaag caagcactga tgaagctcac 1560
cctagagtct tttattcaaa acttcagcgt tctctatagt cccttgaaaa ggcatctcat 1620
tggaagtggc tctgcccagt tcccgtctca gcatttaatc actgaagtga caactgatac 1680
cttttgggaa gtagtccttc aaaaacagga cgttctcctg ctctattacg ctccgtggtg 1740
cggcttctgt ccatccctca atcacatctt catccagcta gctcggaacc tgcccatgga 1800
cacattcact gtggcaagga ttgacgtgtc tcagaatgac cttccttggg aatttatggt 1860
cgatcgtctt cctactgtct tgtttttttcc ctgcaacaga aaggacctaa gtgtgaaata 1920
ccccgaagac gtccccatca cccttccaaa cctgttgagg ttcattttgc atcactcaga 1980
ccctgcttcc agcccccaga atgtggctaa ctctcctacc aaggagtgtc ttcagagcga 2040
ggcagtctta cagcgggggc acatctccca cttggagaga gagatccaga aactgagagc 2100
agaaataagc agcctccagc gagcacaagt gcaggtggag tcccagctct ccagtgcccg 2160
cagagatgag caccggctgc ggcagcagca gcgggccctg gaagagcagc acagcctgct 2220
ccacgcacac agtgagcagc tgcaggccct ctatgagcag aagacacgtg agctgcagga 2280
gctggcccgc aagctgcagg agctggccga tgcctcagaa aacctcctta ccgagaacac 2340
gtggctcaag atcctggtgg cgaccatgga gaggaaactg gagggcaggg atggagctga 2400
aagcctggcg gcccagagag aggtccaccc caagcagcct gagccctcag ccacccccca 2460
gctccctggc agctcccctc cacctgccaa tgtcagcgcc acactggtgt ctgaaaggaa 2520
taaggagaac aggacagact aacttttttaa atgatatgaa gaaatcagag gtgaaaattg 2580
tacattggga atatatttat gcaaatttta ttgaaattta ttgtaaataa agattttctc 2640
agtggtctag aaaaaaaaaa aaaaa 2665

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organizati n
International Bureau

(43) International Publication Date
22 February 2001 (22.02.2001)

PCT

(10) International Publication Number
WO 01/12790 A3

(51) International Patent Classification[7]: C12N 15/61, 9/90, C07K 16/40, C12Q 1/68, A61K 38/52, A01K 67/027, G01N 33/50

(21) International Application Number: PCT/US00/22518

(22) International Filing Date: 16 August 2000 (16.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/149,388 17 August 1999 (17.08.1999) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/149,388 (CIP)
Filed on 17 August 1999 (17.08.1999)

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **BANDMAN, Olga** [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). **LU, Dyung, Aina, M.** [US/US]; 55 Park Belmont Place, San Jose, CA 95136 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **TRAN, Bao** [US/US]; 750 Salberg Avenue, Santa Clara, CA 95051 (US). **HILLMAN, Jennifer, L.** [US/US]; 230 Monroe Drive, #17, Mountain View, CA 94040 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **LAL, Preeti** [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). **TANG, Y., Tom** [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report:
25 October 2001

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/12790 A3

(54) Title: ISOMERASE PROTEINS

(57) Abstract: The invention provides human isomerases (ISOM) and polynucleotides which identify and encode ISOM. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of ISOM.

INTERNATIONAL SEARCH REPORT

Inter ɔnal Application No
PCT/US 00/22518

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/61 C12N9/90 C07K16/40 C12Q1/68 A61K38/52
A01K67/027 G01N33/50

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K C12Q A61K A01K G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ, EMBL, STRAND, MEDLINE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	APEL T W ET AL.: "The ribose 5-phosphate isomerase-encoding gene is located immediately downstream from that encoding murine immunoglobulin kappa" GENE, vol. 156, no. 2, 1995, pages 191-197, XP004042354 Note: 98.4 % aa seq identity of the human RPI protein with SEQ ID NO:2 in 126 aa overlap, 98.6 % nt seq identity with the human RPI gene with SEQ ID NO:10 in 730 nt overlap. abstract figures 1,2 --- -/--	1-19,22, 25-28

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
13 February 2001	15.05.01

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	van de Kamp, M

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Inter onal Application No
PCT/US 00/22518

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL [Online] EMBL; ID AI692341, AC AI692341, 3 June 1999 (1999-06-03) STRAUSBERG R: "wd85g04.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2338422 3', mRNA sequence" XP002160176 Note: 99.5 % nt seq identity with SEQ ID NO:10 in 565 nt overlap. the whole document ---	1-19,22, 25-28
X	DATABASE EMBL [Online] EMBL; ID HS1225161, AC AA425742, 24 May 1997 (1997-05-24) HILLIER L ET AL.: "zw47e12.s1 Soares total fetus Nb2HF8 9w Homo sapiens cDNA clone IMAGE:773230 3' similar to SW:RPIA HUMAN P49247 RIBOSE 5-PHOSPHATE ISOMERASE ;, mRNA sequence" XP002160177 Note: 100.0 % nt seq identity with SEQ ID NO:10 in 516 nt overlap. the whole document ---	1-19,22, 25-28
A	WO 97 43426 A (BRAXTON SCOTT MICHAEL ;INCYTE PHARMA INC (US); MURRY LYNN E (US)) 20 November 1997 (1997-11-20) the whole document ---	
A	US 5 723 311 A (ADAMS MARK D ET AL) 3 March 1998 (1998-03-03) the whole document ---	
A	WO 95 21861 A (MERCK & CO INC ;WIEDERRECHT GREGORY J (US); SEWELL TONYA J (US)) 17 August 1995 (1995-08-17) the whole document -----	

INTERNATIONAL SEARCH REPORT

I. national application No.
PCT/US 00/22518

B x I Observations where certain claims wer found unsearchable (Continuation of it m 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

Remark (1): Although claims 18, 21 and 24 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition. Concerning claims 21 and 24, see also Remark (2).

2. [X] Claims Nos.: 20, 21, 23, 24
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

see FURTHER INFORMATION sheet PCT/ISA/210

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [X] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-28 all partially

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-28 (all partially)

   An isolated polypeptide comprising or consisting of the amino acid sequence of SEQ ID NO:2 or homologs or fragments thereof, an isolated polynucleotide encoding a polypeptide as said, a recombinant polynucleotide containing a polynucleotide as said linked to a promoter, a host cell and a transgenic organism containing a recombinant polynucleotide as said, a method for producing a polypeptide as said, an antibody binding to a polypeptide as said, an isolated polynucleotide comprising a nucleotide sequence of SEQ ID NO:10 or homologs or fragments thereof, methods of detecting a target polynucleotide having a nucleotide sequence of a polynucleotide as said, a pharmaceutical composition comprising a polypeptide as said and the use of a composition as said in medical treatment, methods for screening for a compound that is effective as an agonist or an antagonist of a polypeptide as said, and compositions comprising a agonist or an antagonist as said as well as the use of compositions as said in medical treatment, methods for screening for a compound that binds to a polypeptide as said, that modulates the activity of a polypeptide as said, or that alters the expression of a polynucleotide as said, and a method for assessing the toxicity of a test compound involving the detection of a polynucleotide as said.

2. Claims: 1-28 (all partially)

   As invention 1, but concerning SEQ ID NOs 3 and 11.

3. Claims: 1-28 (all partially)

   As invention 1, but concerning SEQ ID NOs 4 and 12.

4. Claims: 1-28 (all partially)

   As invention 1, but concerning SEQ ID NOs 5 and 13.

5. Claims: 1-28 (all partially)

   As invention 1, but concerning SEQ ID NOs 6 and 14.

6. Claims: 1-28 (all partially)

   As invention 1, but concerning SEQ ID NOs 7 and 15.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

7. Claims: 1-28 (all partially)

As invention 1, but concerning SEQ ID NOs 8 and 16.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 20, 21, 23, 24

Remark (2): Claims 20, 21, 23 and 24 refer to agonists and antagonists of the polypeptide, and to compositions comprising such agonists and antagonists, without giving a true technical characterisation. Moreover, no specific compounds are defined in the application. In consequence, the scope of said claims is ambiguous and vague, and their subject-matter is not sufficiently disclosed and supported (Art. 5 and 6 PCT). No search can be carried out for such purely speculative claims whose wording is, in fact, a mere recitation of the results to be achieved.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inter nal Application No
PCT/US 00/22518

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9743426 A	20-11-1997	US 5798249 A	25-08-1998
		AU 3066197 A	05-12-1997
		US 6001632 A	14-12-1999
US 5723311 A	03-03-1998	NONE	
WO 9521861 A	17-08-1995	US 5457182 A	10-10-1995
		CA 2181803 A	17-08-1995
		EP 0749443 A	27-12-1996
		JP 9509160 T	16-09-1997